UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

PhD. Thesis

On the Automatic Detection of Otolith Features for Fish Species Identification and their Age Estimation

Author:

José Antonio Soria Pérez

Advisor:

Vicenç Parisi Baradad

Advanced Hardware Architectures Group (AHA)

DEPARTMENT OF ENGINEERING ELECTRONICS
UNIVERSITAT POLITÈNICA DE CATALUNYA

Barcelona, October 2012

Acknowledgments

I am grateful to many people and institutions that, directly and indirectly, have made this PhD. thesis possible.

I would like to acknowledge the help and support given by *Universitat Politècnica de Catalunya* (UPC), particularly, the *Departament d'Enginyeria Electrònica* (EEL), the administrative staff and, above all, the *Advanced Hardware Architectures* Group (AHA). I would also like to thank *the Institut de Ciències del Mar* (ICM-CMIMA) for providing the AFORO database resources, without which I could not have developed this thesis.

I am very thankful to my advisor Vicenç Parisi. His personal vision regarding image and signal processing applications has always made things look easier and his support helped me through the difficult periods. I am also thankful to Antoni Lombarte, from CMIMA, for the help he provided in the biological interpretation of results, and to Gabriel Torres who prepared the otolith images.

I would also like to thank the *Institut Français de Recherche pour l'Explotation de la MER* (IFREMER), specially R. Fablet and H. de Pontual for letting me participate in the AFISA project. I gleaned both experience and maturity in addressing real problems from participating in the project.

I am thankful to Joan Cabestany for his useful intuitions and personal comments. I also owe a great deal to my professors and office-mates for the cheery company during so many coffees, lunches and class break talks. In this sense, the most significant contributions have been made by C. Raya, F.J. Ruiz, M. López and R. Ramos. My deepest gratitude for so many talks, discussions, advice and good times in general.

I feel very much indebted to my parents Antonio and Maria. This PhD. thesis is specially dedicated to my mother, for the zealous dedication, entirety and infinite patience she showed during such difficult times.

Vilanova i la Geltrú, October 2012

José Antonio Soria Pérez

Resum

L'eix principal d'aquesta tesi tracta sobre la detecció automàtica d'irregularitats en senyals, tant si s'extreuen de les imatges fotogràfiques com si es capturen de sensors electrònics, així com la seva possible aplicació en la detecció d'estructures morfològiques en otòlits de peixos per identificar espècies, i realitzar una estimació de l'edat en el moment de la seva mort.

Des de la vesant més biològica, els otòlits, que son estructures calcàries que es troben en el sistema auditiu de tots els peixos teleostis, constitueixen un dels elements principals en l'estudi i la gestió de l'ecologia marina. En aquest sentit, l'ús combinat de descriptors de Fourier i l'anàlisi de components es el primer pas i la clau per caracteritzar la seva morfologia i identificar espècies marines. No obstant, una de les limitacions principals d'aquest sistema de representació consisteix en la interpretació limitada de les irregularitats que pot desenvolupar, així com l'ús que es realitza dels coeficients en tasques de classificació, els quals, acostumen a ser seleccionats manualment tant pel que respecta a la quantitat com la seva importància.

La detecció automàtica d'irregularitats en senyals, així com la seva interpretació, es va tractar per primera vegada sota el marc del *Best-Basis paradigm*. En aquest sentit, l'algorisme *Local Discriminant Bases* (LDB) de N. Saito es basa en la *Transformada Wavelet Discreta* (DWT) per descriure el posicionament de característiques dintre de l'espai temporal-freqüencial, i en una mesura discriminant basada en l'energia per guiar la cerca automàtica de característiques dintre d'aquest domini. Propostes més recents basades en funcions de densitat han tractat de superar les limitacions de les mesures d'energia amb un èxit relatiu. No obstant, encara s'han de desenvolupar noves estratègies que siguin més consistents amb la capacitat real de classificació i ofereixin més generalització al reduir la dimensió de les dades d'entrada.

La proposta d'aquest treball es centra en un nou marc per senyals unidimensionals. Una de las conclusions principals que s'extreu es que aquesta generalització passa per establir un marc de mesures acotades on els valors reflecteixin la densitat on cap classe es solapa. Això condiciona bastant el procés de selecció de característiques i la mida del vector necessari per identificar les classes correctament, que s'han d'establir no només en base a valors discriminants globals si no també en informació complementària sobre la disposició de les mostres en el domini.

Les noves eines s'han utilitzat en diferents estudis d'espècies de lluç, on s'han obtingut bons resultats d'identificació. No obstant, l'aportació principal consisteix en la interpretació que l'eina extreu de les característiques seleccionades, i que inclou l'estructura de les irregularitats, la seva posició temporal-freqüencial, extensió en l'eix i grau de rellevància, el qual, es ressalta automàticament sobre les mateixa imatge o

senyal.

En quan a l'àmbit de determinació de l'edat, s'ha plantejat una nova estratègia de demodulació de senyals per compensar l'efecte del creixement no lineal en els perfils d'intensitat. Tot i que inicialment aquesta tècnica desenvolupa un procés d'optimització capaç d'adaptar-se automàticament al creixement individual de cada peix, els resultats amb el LDB suggereixen estudiar l'efecte de les condicions lumíniques sobre els otòlits amb la finalitat de dissenyar algorismes que redueixin la variació del contrast de les imatges més fiablement.

Mentrestant s'ha plantejat una nova teoria per realitzar estimacions d'edat en peixos en base als otòlits. Aquesta teoria suggereix que si la corba de creixement és coneguda, el període regular dels anells en el perfil d'intensitat demodulat està relacionat amb la longitud total de radi d'on s'agafa el perfil original. Per tant, si la periodicitat es pot mesurar, es possible conèixer l'edat exacta del peix sense usar extractors de característiques o classificadors, la qual cosa tindria implicacions importants en l'ús de recursos computacionals i en les tècniques actuals d'estimació de l'edat.

Resumen

El eje principal de esta tesis trata sobre la detección automática de singularidades en señales, tanto si se extraen de imágenes fotográficas como si se capturan de sensores electrónicos, así como su posible aplicación en la detección de estructuras morfológicas en otolitos de peces para identificar especies, y realizar una estimación de la edad en el momento de su muerte.

Desde una vertiente más biológica, los otolitos, que son estructuras calcáreas alojadas en el sistema auditivo de todos los peces teleósteos, constituyen uno de los elementos principales en el estudio y la gestión de la ecología marina. En este sentido, el uso combinado de descriptores de Fourier y el análisis de componentes es el primer paso y la clave para caracterizar su morfología e identificar especies marinas. Sin embargo, una de las limitaciones principales de este sistema de representación subyace en la interpretación limitada que se puede obtener de las irregularidades, así como el uso que se hace de los coeficientes en tareas de clasificación que, por lo general, acostumbra a seleccionarse manualmente tanto por lo que respecta a la cantidad y a su importancia.

La detección automática de irregularidades en señales, y su interpretación, se abordó por primera bajo el marco del *Best-Basis paradigm*. En este sentido, el algoritmo *Local Discriminant Bases* (LDB) de N. Saito utiliza la *Transformada Wavelet Discreta* (DWT) para describir el posicionamiento de características en el espacio tiempo-frecuencia, y una medida discriminante basada en la energía para guiar la búsqueda automática de características en dicho dominio. Propuestas recientes basadas en funciones de densidad han tratado de superar las limitaciones que presentaban las medidas de energía con un éxito relativo. No obstante, todavía están por desarrollar nuevas estrategias más consistentes con la capacidad real de clasificación y que ofrezcan mayor generalización al reducir la dimensión de los datos de entrada.

La propuesta de este trabajo se centra en un nuevo marco para señales unidimensionales. Una conclusión principal que se extrae es que dicha generalización pasa por un marco de medidas de valores acotados que reflejen la densidad donde las clases no se solapan. Esto condiciona severamente el proceso de selección de características y el tamaño del vector necesario para identificar las clases correctamente, que se ha de establecer no sólo en base a valores discriminantes globales sino también en la información complementaria sobre la disposición de las muestras en el dominio.

Las nuevas herramientas han sido utilizadas en el estudio biológico de diferentes especies de merluza, donde se han conseguido buenos resultados de identificación. No obstante, la contribución principal subyace en la interpretación que dicha herramienta hace de las características seleccionadas, y que incluye la estructura de las irregularidades, su posición temporal-frecuencial, extensión en el eje y grado de relevancia, el

cual, se resalta automáticamente sobre la misma imagen o señal.

Por lo que respecta a la determinación de la edad, se ha planteado una nueva estrategia de demodulación para compensar el efecto del crecimiento no lineal en los perfiles de intensidad. Inicialmente, aunque el método implementa un proceso de optimización capaz de adaptarse al crecimiento individual de cada pez automáticamente, resultados preliminares obtenidos con técnicas basadas en el LDB sugieren estudiar el efecto de las condiciones lumínicas sobre los otolitos con el fin de diseñar algoritmos que reduzcan la variación del contraste de la imagen más fiablemente.

Mientras tanto, se ha planteado una nueva teoría para estimar la edad de los peces en base a otolitos. Esta teoría sugiere que si la curva de crecimiento real del pez se conoce, el período regular de los anillos en el perfil demodulado está relacionado con la longitud total del radio donde se extrae el perfil original. Por tanto, si dicha periodicidad es medible, es posible determinar la edad exacta sin necesidad de utilizar extractores de características o clasificadores, lo cual tendría implicaciones importantes en el uso de recursos computacionales y en las técnicas actuales de estimación de la edad.

Abstract

This thesis deals with the automatic detection of features in signals, either extracted from photographs or captured by means of electronic sensors, and its potential application in the detection of morphological structures in fish otoliths so as to identify species and estimate their age at death.

From a more biological perspective, otoliths, which are calcified structures located in the auditory system of all teleostean fish, constitute one of the main elements employed in the study and management of marine ecology. In this sense, the application of Fourier descriptors to otolith images, combined with component analysis, is habitually a first and key step towards characterizing their morphology and identifying fish species. However, some of the main limitations arise from the poor interpretation that is sometimes obtained with this representation and the use that is made of the coefficients, as they are usually selected manually for classification purposes, both in quantity and representativity.

The automatic detection of irregularities in signals, and their interpretation, was first addressed in the so-called *Best-Basis paradigm*. In this sense, Saito's *Local Discriminant Bases* algorithm (LDB) uses the *Discrete Wavelet Packet Transform* (DPWT) as the main descriptive tool for positioning the irregularities in the time-frequency space, and an energy-based discriminant measure to guide the automatic search of relevant features in this domain. Current density-based proposals have tried to overcome the limitations of energy-based functions but with relatively little success. However, other measuring strategies which are more consistent with true classification capability and which provide generalization while at the same time reduce the dimensionality of input features, are yet to be developed.

The proposal of this work focuses on a new framework for one-dimensional signals. An important conclusion extracted therein is that such generalization involves a measurement system of bounded values representing the density where no class overlaps. This acutely determines the feature selection process and the vector size that is required for proper class identification, which must be implemented not only based on global discriminant values but also on complementary information regarding the provision of samples in the domain.

These new tools have been used in the biological study of different species of hake, and have yield good classification results. However, a major contribution lies in the further information the tool is able to interpret from the selected features, including the shape of irregularities, their time-frequency position, extension support and degree of importance, which is highlighted automatically on the same images or signals.

As for aging applications, a new demodulation strategy for compensating the nonlinear growth effect on the intensity profile has been developed. Although the new method develops an optimization process which can, in principle, adapt automatically to the specific growth of individual specimens, preliminary results with LDB-based techniques suggest that the effect of lightning conditions on the otoliths should be studied to design algorithms which reliably reduce image contrast variation.

In the meantime, a new theoretical framework for otolith-based fish age estimation has been presented. This theory suggests that if the true fish growth curve is known, the regular periodicity of age structures in the demodulated profile is related to the radial length the original profile is extracted from. Therefore, if this periodicity can be measured, it is possible to infer the exact age of the fish and thus omit feature extractors and classifiers. This could have important implications in the use of computational resources and current aging approaches.

Contents

Acknowledgements i

Abstract iii

Contents ix

List of Figures xv

List of Tables xxii

I Otoliths and Applications xxv

1 Introduction 1

1.1 The Importance of Otoliths 1

1.2 The Use of Otoliths in Marine Applications 5

1.2.1 Identification of Fish Species 5

1.2.2 Aging Applications 6

1.3 On Pattern Recognition Systems 8

1.4 Motivations and Scope of the Thesis 11

1.5 Overview of this Document 13

2 Acquisition of Otolith Data 15

2.1 Introduction 15

2.2 Basic Otolith Preparation 16

2.2.1 Direct Observation 16

2.2.2 Embedding for Incident-light Observations 16

2.2.3 Manual Validation of Age Readings 20

2.2.4 Image Quality Issues 22

2.3 Contouring by Segmentation . . . 23
2.4 Extracting Age Signals . . . 25
2.4.1 Nucleus Detection . . . 26
2.4.1.1 Maxima Removal and Grain Filtering . . . 26
2.4.1.2 Principal Axis Proximity . . . 27
2.4.1.3 Geometrical Statistical Selection . . . 28
2.4.2 Extraction and Preprocessing of Age Structures . . . 31
2.4.2.1 One-dimensional Intensity Profiles . . . 31
2.4.2.2 Supervised Growth Demodulation (SGD) . . . 32
2.4.2.3 Contrast cancellation . . . 36
2.4.2.4 Peak-Based Representation (PB) . . . 37
2.5 Discussion . . . 38

II On Feature Extraction 41

3 Feature Extraction 43
3.1 Introduction . . . 43
3.1.1 Why Wavelets . . . 44
3.1.2 Overview of this Chapter . . . 45
3.2 On Description Methods for Feature Extraction . . . 45
3.2.1 Fourier Descriptors . . . 47
3.2.2 STFT. The Short-Time Fourier Transform . . . 48
3.2.3 Wavelet Transforms . . . 50
3.2.3.1 Discrete Wavelet Packet Transform (DWPT) . . . 57
3.2.4 Block Transforms . . . 59
3.2.4.1 Block Orthogonal Basis . . . 59
3.2.4.2 Cosine and Sine Transforms . . . 60
3.2.4.3 Block Cosine Transform . . . 63
3.2.4.4 Lapped Projectors . . . 64
3.2.4.5 Local Cosine Basis . . . 65
3.2.4.6 DCPT. Cosinus Packet Transform Computation . . . 66
3.3 On Feature Extraction Methods . . . 68
3.3.1 Best-Basis Selection (BB) . . . 68
Algorithm 3.1: Best Basis . . . 69
3.3.1.1 Dimensionality Reduction . . . 70

3.3.1.2 Best-Basis Selection from a Library of Orthonormal Bases . 70
3.3.1.3 Karhunen-Loève Basis and Joint Best-Basis 71
3.3.2 Linear Discriminant Analysis (LDA) 72
3.3.3 The Local Discriminant Basis (LDB) 73
3.3.3.1 Divide-and-Conquer 74
Algorithm 3.2: Local Discriminant Basis 74
3.3.3.2 A Library of Local Discriminant Bases 75
3.4 Feature Representation Tools 76
3.4.1 Time-frequency Support 77
3.4.2 Visualization of Data in $\mathbb{R}^N$ 78
3.5 Conclusions 79

4 Classifiers **81**
4.1 Introduction 81
4.2 Problem Formulation 82
4.2.1 Estimation of Accuracy 82
4.2.2 The Bayes Rule 85
4.3 Bayes Classifiers 86
4.3.1 Joint and Naive Classifiers 86
4.3.2 LDA-Based Classifiers 87
4.3.3 k-Nearest Neighbor (k-NN) 88
4.4 Learning Vector Quantization (LVQ) 89
4.4.1 LVQ algorithms 90
4.4.1.1 The LVQ1 90
4.4.1.2 2nd Version (the LVQ2) 91
4.4.1.3 3rd Version (the LVQ3) 91
4.4.1.4 Differences between LVQ1, LVQ2 and LVQ3 92
4.4.1.5 Optimizing the Learning Rate 92
4.4.2 General Considerations 93
4.4.2.1 Initialization of Codebook Vectors 93
4.4.2.2 Learning and Stopping Rule 94
4.5 Conclusions 94

III Contributions **97**

5 On LDB-based Pattern Recognition **99**

5.1 Introduction 99
5.2 1st. Proposition: Size, Rotation and Translation Normalization 101
5.3 Density Estimation for Local Feature Description 103
5.3.1 Histograms 103
5.3.1.1 Bin Width Selection by Cross-validation 105
5.4 2nd. Proposition: Density Local Discriminant Basis 107
Hypothesis 1. Normalized density 107
Hypothesis 2. Discrimination-Accuracy Correlation 108
5.4.1 Towards the Standardization of δ 109
5.4.2 Node Selection Combined with Top-down Search 110
Algorithm 5.1: Density Local Discriminant Basis 111
5.5 3th. Proposition: Automatic Growth Demodulation (AGD) 112
Definition 5.1: Optimal Fish Growth Estimation 113
5.6 Conclusions 114

IV Results and Conclusions **115**

6 Fish Identification and Age Estimation Results **117**
6.1 Introduction 117
6.2 Application by Fish Identification 120
6.2.1 Analysis of *M. merluccius* and *G. morhua* 120
6.2.1.1 Results by Fourier Description 121
6.2.1.2 Results by PCA and LDA 121
6.2.1.3 LDB and DLDB analysis 123
6.2.1.4 Classification of *M. merluccius* and *G. morhua* 127
6.2.2 Results by Intra-specific and Inter-specific Analysis of *Merluccius* Populations 130
6.2.2.1 Inter-specific Experimentation 132
6.2.2.2 Intra-specific experimentation 137
6.2.3 Discussion of Results 143
6.2.3.1 Feature Selection Methodology 143
6.2.3.2 Ecological Significance of *Merluccius* Populations 146
6.3 Application by Age Estimation 147
6.3.1 Manual vs. Automatic Contrast Cancellation and Signal Demodulation 148
6.3.2 Discussion 151

6.4 Conclusions . . . 152
6.4.1 Feature Extraction . . . 152
6.4.2 Signal Demodulation . . . 154

7 Final Remarks and Further Development **157**
7.1 General Conclusions . . . 157
7.2 Feature Extraction Methodology . . . 159
7.2.1 Input Data . . . 159
7.2.2 Filter Design . . . 159
7.2.3 Translation Dependence . . . 160
7.2.4 Considerations for Multivariate Densities . . . 160
7.2.5 Search Strategies for C Classes . . . 161
7.2.6 Future Proposal . . . 164
Algorithm 7.1: The DLDB2 . . . 164
7.3 Age Estimation . . . 165
7.3.1 On Automatic Contrast Cancellation for Aging Technology . . . 166
7.3.2 About the Use of Classifiers for Aging Purposes . . . 169
Hypothesis 3: Definition of Fish Age . . . 169
7.4 Final Remark . . . 169

V Appendices **171**

A Wavelet Design **173**
A.1 Introduction . . . 173
A.2 Multiresolution Signal Processing . . . 173
A.2.1 Discrete Signals . . . 175
A.2.2 FIR Filter Banks and Compactly Supported Wavelets . . . 177
A.2.2.1 Bases of Orthonormal Wavelets Constructed from Filter Banks 180
A.2.3 General FIR Perfect Reconstruction Filter Banks . . . 182
A.2.3.1 Orthogonal or Paraunitary Filter Banks . . . 184
A.2.3.2 Biorthogonal or General Perfect Reconstruction Filter Banks 186
A.3 Filter Design . . . 190
A.3.1 Some Notes on Biorthogonal Filter Development . . . 192
A.3.1.1 Linear Phase Filter Design with a Maximum Number of Zeros at π . . . 193
A.3.1.2 *Diophantine* Equations and Complementary Filters . . . 194

A.3.1.3 Discussion 196

B CWT Computation with the DWT **197**

B.1 Finner Sampling in Scale 197

B.2 Finner Sampling in Time: The "à trous" Algorithm 198

C Multivariate Density Algorithms **201**

C.1 Introduction 201

C.2 Kernel Density Estimation (KDE) 202

C.3 The Nearest Neighbor Approach (k-NN) 206

Bibliography **209**

Subject Index **229**

List of Figures

1.1 Illustration of two common techniques for inter-specific and intra-specific analysis of fish specimens: a) *Landmarks* (from Thompson 1917); b) *Outlines* 2

1.2 Right inner ear of a *Merluccius capensis* specimen. All sensing *maculae* and innervations of the nerves in the epithelium can be observed. as - *asteriscus* otolith; lp - *lapillus* otolith; s, *saccula*; sc - semicircular channels; sg, *sagitta* otolith (from Lombarte 1990). 3

1.3 Right *sagitta* otolith from a *M. capensis* specimen: ca - *caudal* area ; cc - *collicum caudal*; cl - *collum*; co - *collicum-ostial*; dcm - *dorso-caudal* margin; drm - *dorso-rostral* margin; ra - *rostral* area; sa - *sulcus acusticus*; vc - ventral *caudal*; vr - ventral *rostral* (from Lombarte 1990) 3

1.4 Conventional terminology used to describe the shape of otolith outlines (from Tuset et al. 2008). 4

1.5 Example of manual age estimation for a 5-year cod otolith (source: IBACS European project). 7

1.6 Ring extraction example in feature-based methods. Sector selection and unfolding are common processes for both groups of techniques (feature-based and filter-based methods) and require manual implementation. An optimal contrast variation at the otolith border is critical for estimating the age. 7

1.7 Flow diagram of generalized pattern recognition systems (from Osuna 2005). Feature choice and model validation are the main theme of this thesis. 9

2.1 Preparation of the catalyst. The volume required depends on the full amount of polyester resin required according to Table 2.2 (from McCurdy et al. 2002) 18

2.2 Example of otolith embedding in an elastomer mould. a) A polymerized resin layer (yellow arrow) is run on the bottom of each location before the otolith is deposited on the bottom. b) The otolith is oriented and embedded in a second layer. c) It should be made sure that the resin fills the mould completely. In this case the resin does not reach the edge so that it can be turned over to drive out the air bubbles. d) Otolith definitively embedded (red arrow) (from McCurdy et al. 2002) 20

2.3 The same plaice otolith image captured for aging tasks. (a) *reflected-light* and (b) *transmitted-light*, respectively. Gray-scale levels corresponding to translucent and opaque regions are inverted. 23

2.4 Example of image segmentation using the Otsu method: (a) Original gray-scale image; (b) Binarized image. 24

2.5 Nucleus detection algorithm scheme (from Cao and Fablet 2006). A dotted line is used to indicate the 2nd pass by which the empirical probability $E_2(S)$ is calculated and the final shape for the nucleus (N) determined. 27

2.6 Proximity to principal axis (from Cao and Fablet 2006). The black-filled shape is in $\mathcal{E}$ since in any direction orthogonal to $\triangle$, it is contained in the shape closest to $\triangle$ (the white one). For the same reason, the white shape belongs to $\mathcal{E}$. On the contrary, one can find a normal to $\triangle$ such that the gray-filled shape is not closest to $\triangle$ (since the white is closer). Therefore, it does not belong to $\mathcal{E}$. 28

2.7 Illustration of the template-based approach aimed at extracting a 1D signal representation of the image content within a region of interest under transmitted-light (from Fablet 2006). Yearly translucent rings are indicated by markers (left). The relevant points of the region of interest are represented (right). 31

2.8 Illustration of the unwarping (or demodulation) operation. The blue trace represents the original intensity profile, i. The irregular positions in l-domain are changed to a new t-domain by means of the inverse of the otolith growth function, $t = v_{-1}(l)$, in order to obtain another profile, i_{DM}, of regular growth rings 33

2.9 illustration of the peak-based representation for the otolith image represented in Figure 2.7 (from Fablet 2006). a) Signal $i(l)$ with its extracted maxima positions (green markers), b) Associated peak-based representation $i_{PB}(l)$. 38

3.1 Examples of basis functions available for feature description. From left to right: the Daubechies 4 wavelet packet basis, the Haar basis and the discrete sine basis. Top row: Translated basis functions. Bottom row: Basis functions with different frequencies. Horizontal axes indicate time in this figure. 46

3.2 Heisenberg boxes of two windowed Fourier atoms $\phi_{u,\xi}$ and $\phi_{v,\gamma}$ (from Mallat 1990). 50
3.3 Heisenberg boxes of two wavelets. Smaller scales decrease the time spread but increase the frequency support, which is shifted towards the higher frequencies (from Mallat 1990). 52
3.4 A decomposition of V_0 into mutually orthogonal spaces using the wavelet transform (with $J = 3$). The symbols in bold font represent the subspaces kept intact by the discrete wavelet transform. 53
3.5 Fast discrete wavelet transform computation (from Mallat 1990); (a) A fast wavelet transform is computed with a cascade of filterings with $\overline{\boldsymbol{h}}_{0'}$ and $\overline{\boldsymbol{h}}_1$ followed by a factor 2 sub-sampling; (b) A fast inverse wavelet transform reconstructs progressively each $\boldsymbol{a}_j$ by inserting zeros between samples of $\boldsymbol{a}_{j+1}$ and $\boldsymbol{d}_{j+1}$, filtering with $\boldsymbol{g}_0$ and $\boldsymbol{g}_1$, and adding both sequences. 55
3.6 Variance translation in the computation of DWT (from Mallat 1990). If $x_\tau(t) = x(t-\tau)$, uniformly sampling $C_{2^j,u}$ and $C_{2^j,u-\tau}$ at $u = n2^j$ may yield very different values if $\tau \neq n2^j$ 56
3.7 A decomposition of $W_{0,0}$ into the tree-structured subspaces using the wavelet packet transform (with $J = 3$) 57
3.8 Fast DWPT algorithm (from Mallat 1990); (a) A fast wavelet transform is computed with a cascade of filterings with $\overline{h}$ and $\overline{g}$ followed by a factor 2 sub-sampling; (b) A fast inverse wavelet transform reconstructs progressively each a_j by inserting zeros between samples of a_{j+1} and d_{j+1}, filtering and adding both sequences. . . 58
3.9 Example of an admissible tree for a sequence of N=8, decomposed up to a maximum scale of J=3. The selection of nodes is carried out in a way such that the full time and frequency axis is covered with N basis functions of different time-scale resolutions. 59
3.10 Signal extensions of the main cosinus transforms (from Mallat 1990); a) Extension for the DCT-II; b) Extension for the DCT-IV. A cosine II extends $x[n]$ into a signal of period 2 which is symmetric around 0. The extension in a cosinus-IV, however, is asymmetric with respect to N and of period 4. 63
3.11 Each window g_p has a support $[a_p - \eta, a_{p+1} + \eta]$ with an increasing profile and decreasing profile over $[a_p - \eta, a_p + \eta]$ and $[a_{p+1} - \eta, a_{p+1} + \eta]$. 64
3.12 An admissible binary tree of local cosine spaces divides the time axis in windows of dyadic lengths. 67
3.13 Flow diagram for the Local Discriminant Basis feature selection algorithm. The Best-Basis paradigm involves *description* and *selection* of features. Proposals for the three first blocks are considered in this thesis (see chapter 5). 75

5.1 Limitations of the use of the energy magnitude in the evaluation of discrimination power. f_a and f_b are the density distributions of two well separated classes, whereas f_c and f_d overlap each other, despite having much more energy 100

5.2 Detailed diagram of the proposed feature extraction scheme in this thesis. Our proposals involve methods for both fish identification and aging applications (blocks highlighted in red) and extend the general diagram depicted in Fig. 3.13. The new LDB implements the DWPT, density estimation, discriminant analysis and a divide-and-conquer algorithm. Dimensionality reduction is developed manually. 101

5.3 Normalization by the semi-major axis of the first Fourier harmonic ellipse. Distance (OA) denotes the radial of this semi-major axis. Normalization is performed rotating the otolith contour by ψ and shifting all coordinate point by the amount of λ 102

6.1 World distribution of the ten *Merluccius* species considered in this work (based on Lloris et al. 2005). a) North American Pacific and Atlantic (Na1: *M. productus*, Na2: *M. bilinearis*). b) Europe (Eu1: *M. merluccius*). c) South American Pacific and Atlantic (Sa1: *M. gayi*, Sa2: *M. australis*). d) Africa (Af1: *M. senegalensis*, Af2: *M.polli*, Af3: *M. capensis*, Af4: *M. paradoxus*). 119

6.2 Normalized otolith contour examples from *Merluccius merluccius* (blue) and *Gadus morhua* (green). a) Polar representation b) Clockwise representation of the radials. The coordinates have been resampled to 256 points in order to use the 1D version of the DLDB 120

6.3 Fourier analysis of *Merluccius merluccius* (blue) and *Gadus morhua* (green) specimens. a) the fourteen lowest frequency components (offset component f_0, not included). b) Radial reconstruction using $f_0 - f_8$ 121

6.4 PCA results for *M. merluccius* (blue) and *G. morhua* (green). The distribution of STD and DFT coefficients is represented by means of scatter maps (left and right column, respectively). Results are presented for features a-b) $f_1 - f_2$; c-d) $f_1 - f_3$ and e-f) the best 10 PCA coefficients (represented in $\mathbb{R}^2$ by means of the Sammon map dimension reduction technique) 122

6.5 Admissible trees for the *M. merluccius* - *G. morhua* experiment using the *reverse biorthogonal 3.1* wavelets. Best nodes are shown in black and the best five coefficients are shown in red for a) the DLDB, and b) the LDB. 124

6.6 Waveform patterns for the five basis functions selected by the DLDB using the rbior3.1 dictionary. a) These representations correspond to the atoms in table 6.3. b) Their frequency spectrum is calculated by means of DFT. 126

6.7 Detection of differences between cod (green) and hake (blue) at the otolith contour. a) LDB, b) DLDB. Maximum discrimination is obtained for the rbior3.1 wavelets at the dorso-caudal and dorso-rostral margins. 126
6.8 Scatter maps for cod (green) and hake (blue) features using DLDB. Representations are provided for a) the best two coefficients and, b) the best five coefficients (represented in $\mathbb{R}^2$ by means of the Sammon map dimension reduction technique) 127
6.9 Matching the performance of LDB (left column) and DLDB (right column) with classification results. Discrimination values and identification results for both LVQ and k-NN are provided for all 256 selected coefficients of the rbior3.1 wavelet dictionary, a) and b), and all $\mathfrak{D}$ dictionaries, c) and d), by decreasing importance. 129
6.10 Location of otolith differences in inter-specific *M. albidus* (blue) - *M. bilinearis* (green), north-west Atlantic, USA. Feature location is expressed in terms of the momentum, u, and its variance, σ_u, over the contour trace. The colored hot bar indicates the significance of discriminant measures. 133
6.11 Location of otolith differences in inter-specific *M. capensis* (blue) - *M. paradoxus* (green), south-east Atlantic, Namibia. 135
6.12 Location of otolith differences in inter-specific *M. polli* (blue)- *M. senegalensis* (green), central-east Atlantic, Senegal. 136
6.13 Location of otolith differences in inter-specific *M. gayi* (blue)- *M. australis* (green), south-east Pacific. 138
6.14 Location of otolith differences in intra-specific *M. merluccius*, Mediterranean (blue) and north-east Atlantic (green). 140
6.15 Location of otolith differences in intra-specific *M. bilinearis*, north-west Atlantic, USA (blue) and Canada (green). 141
6.16 Location of otolith differences in intra-specific *M. gayi*, south-west Pacific, Chile (blue) and Peru (green). 142
6.17 Location of otolith differences in intra-specific *M. productus*, north-west Pacific, USA (blue) and Canada (green). 144
6.18 Example of automatic detection of the first year mark by means of AGD. The otolith section correspond to a 3 year-old cod specimen. The one-dimensional intensity signal has been extracted from the nucleus to the contour border (right to left) following the green trace. Age information is provided by the expert (red dots) in order to compare the detection of the first year period (red circle). . . . 147
6.19 Extracted intensity profile from the previous cod otolith section in Fig. 6.18. As indicated by the expert (see red marks), year increments don't have a regular rate because of the non-linear fish growth. 148

6.20 Supervised vs. Automatic Growth Demodulation results. Original, manual and automatically demodulated signals (red, blue and green traces, respectively) are represented together with the histogram of autocorrelation lags (green stems). The bin with the majority of elements sets the first year period in the demodulated signal (first blue mark) and establishes an annual rate reference for the remaining year marks. 149
6.21 Plaice otolith samples used in our experiments for different age classes. (a) two years; (b) three years; (c) five years; (d) six years or more. 150
6.22 Some otolith examples from different taxonomic value. a1) *Trachurus mediterraneus;* a2) *Trisopterus minutus;* a3) *Coris julis*; a4) *Umbrina canariensis* and a5) *Scomber colias.* 154

7.1 Naive sequential feature selection example considering independent features. Only the pair combination $\{x_1, x_4\}$ can obtain full class separation 162
7.2 Filter-based FSS and Wrapper-based FSS. The objective function of wrapper techniques are based on the predictive accuracy, which is estimated by statistical resampling or cross-validation methods (from Osuna 2005) 163
7.3 (a) Reflected-light. (b) Refracted light 167
7.4 (a) Specular reflection. (b) Diffuse reflection (from Hecht 2000). 167
7.5 Geometric models for surface reconstruction purposes. (a) central projection; (b) orthogonal parallel projection. (from Klette et al. 1998) 168

A.1 Decomposition of V_{-1} into V_0 using multirate filters, and recombination to achieve perfect reconstruction. $\boldsymbol{H}_0^*\boldsymbol{H}_0\boldsymbol{x}$ is the projection of the signal onto V_0 and $\boldsymbol{H}_1^*\boldsymbol{H}_1\boldsymbol{x}$ is the projection onto W_0. 177
A.2 Discrete wavelet transform on sequences. The half-band low-pass and high-pass filters are $h_0[n]$ and $h_1[n]$, respectively, and $2\downarrow$ means sub-sampling by 2 (dropping odd-indexed samples). 177
A.3 Scaling function generated by using (A.23) for $h_0 = [1, \alpha, \alpha, 1]$ and $\alpha \in \{-3, 3\}$. 179

B.1 Dyadic wavelet transform; (a) The coefficients are computed by cascading convolutions with dilated filters $\overline{\boldsymbol{h}}_0^{(j)}$ and $\overline{\boldsymbol{h}}_1^{(j)}$; (b) The original signal is reconstructed through convolutions with $\boldsymbol{g}_0^{(j)}$ and $\boldsymbol{g}_1^{(j)}$. A multiplication by $1/2$ is necessary to recover subsequent finer scales in $\boldsymbol{a}_j$. 199

C.1 The frequency polygon in a typical bin, $h\left(m-\frac{1}{2}\right) \leq x < h\left(m+\frac{1}{2}\right)$, which is derived from two adjacent histogram bins. 202

C.2 Example of a Gaussian kernel estimate showing the individual kernels for each data point from (Osuna 2005). The kernel function determines the shapes of the bumps, and the "smoothing" parameter, h, determines their width. 205

List of Tables

2.1 Standard mounting materials (from Mosegaard et al. 1998) 17

2.2 Simple dosage of polyester resin (in volume and/or weight) and catalyst (in drops) depending on their ratio. The greater the quantity of catalyst the higher the speed of setting (from Mosegaard et al. 1998) 18

3.1 Summary of the correspondences between the conventional concepts and the new concepts based on the Best-Basis paradigm (or "library" of bases) reviewed or discussed in this chapter. 79

6.1 Results by LDA analysis 123

6.2 Evaluation of different dictionaries with DLDB and LDB. These include the *Daubechies*, *Symlets*, *Coiflets*, *Biorthogonal* and *Reverse Biorthogonal*. Measure values are provided considering the five best features in each dictionary. The best results are in bold format 124

6.3 Information on the location and frequency properties of selected features. The centering position of the atom (and its variability) is in radians. Discrete frequency values have been normalized using $2\xi/N$ 125

6.4 Classification results of *M. merluccius* and *G. morhua*. Identification accuracy is specified in %. The standard deviation calculated by means of the 'hold-out' method is also specified (in brackets). 128

6.5 Available *Merluccius* material for the comparison studies developed in this chapter. Source: AFORO database (http://aforo.cmima.csic.es) and Torres (1997) . 130

6.6 Expected relation between normalized discrimination value and classification accuracy for C classes. 131

6.7 Feature extraction and identification results in inter-specific *M. albidus* - *M. bilinearis*, north-west Atlantic, USA. Accuracy and discrimination values ($\hat{R}^{cv}$ and δ) are specified for each of the five basis function ψ, indexed by the triplet (j, k, l). Locating positions (u) are provided in radians, and the normalization $2\eta/N$ has been used in the discrete frequency. * The additional column shows results for the vector, $\{f_{(1)}, \ldots, f_{(5)}\}$, in $\mathbb{R}^5$. 133

6.8 Feature extraction and identification results in inter-specific *M. capensis* - *M. paradoxus*, south-east Atlantic, Namibia. 135

6.9 Feature extraction and identification results in inter-specific *M. polli* - *M. senegalensis*, central-east Atlantic, Senegal. 136

6.10 Feature extraction and identification results in inter-specific *M. gayi* - *M. australis*, south-east Pacific. 138

6.11 Feature extraction and identification results in intra-specific *M. merluccius*, Mediterranean and north-east Atlantic. 140

6.12 Feature extraction and identification results in intra-specific *M. bilinearis*, north-west Atlantic, USA and Canada. 141

6.13 Feature extraction and identification results in intra-specific *M. gayi*, south-west Pacific, Chile and Peru. 142

6.14 Feature extraction and identification results in intra-specific *M. productus*, north-west Pacific, USA and Canada. 144

6.15 Aging precisions values combining DLDB with the four preprocessing methods: MCE, SGD, AGF, AGD. All four methods are considered independently and values are expressed in terms of agreement to expert estimations. 151

6.16 Fish aging results combining contrast cancellation and growth demodulation. Classification results are provided in terms of expert readings and their known age precision error, $\hat{R}^* = APE \times \hat{R}^{cv}(r)$, in terms of expert agreement which is situated between 85% and 95% for plaice otoliths. 151

7.1 A brief summary on common search strategies 166

C.1 Some common an unusual *Kernels*. All kernels are supported on [-1,1] unless noted otherwise 204

Part I

Otoliths and Applications

Chapter 1

Introduction

1.1 The Importance of Otoliths

The morphological description, analysis and classification of geometrical shapes have become essential tasks in taxonomic studies in biology. These studies can have many different purposes: comparing or identifying species (inter-specific analysis), detecting geographical relationships between similar species (intra-specific analysis) or studying allometric variability in growth of biological species, among other applications (MacLeod 2007).

The use of computer-based image processing tools has played an important role and, over recent decades, has lead to significant advances in the knowledge of fish species (Chesmore 2007). For example, in marine applications two major approaches of shape analysis are common tools for the inter-specific and intra-specific analysis of fish worldwide: *landmarks* and *outlines*.

In general, landmarks (Fig. 1.1a) are useful when studying the variations of homologous points in different species. The goal is to deform the fish contour shape by using some type of axis transformation so that the new shape characteristics can be associated to other species. Although it is an elegant idea, one of the main drawbacks is that the methodology requires additional work to prove that a deterministic relationship exists between the species under consideration (Bookstein 1984, Rohlf and Marcus 1993). On the other hand, the use of outline techniques (Fig. 1.1b) has increased since otoliths became the principal object of analysis in biological studies (Messieh et al. 1989, Lombarte and Castellón 1991, Campana and Casselman 1993). These techniques are based on the simple idea of 'information extraction', where the goal is to characterize the structures of the fish contour shape by means of descriptors so that an analytical process can determine which of them are relevant for explaining the differences of class species (Bookstein et al. 1982, Bookstein 1991).

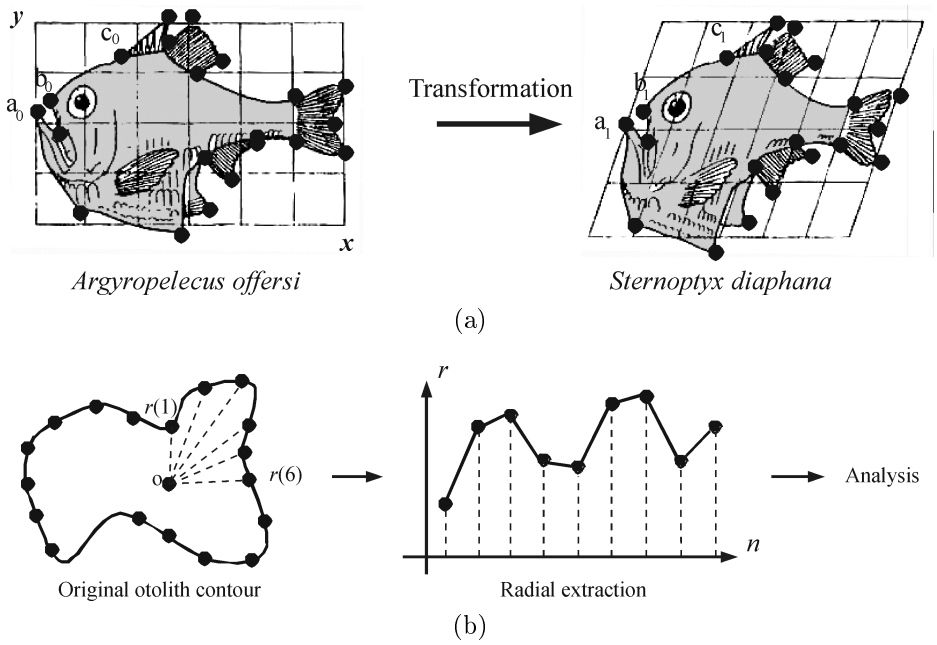


Figure 1.1: Illustration of two common techniques for inter-specific and intra-specific analysis of fish specimens: a) *Landmarks* (from Thompson 1917); b) *Outlines*

Otoliths are calcified structures located in the labyrinthic cavities of the inner ear of all teleostean fish (Blacker 1995, Norman and Greenwood 1975, Platt and Popper 1981). Each ear contains a triplet: the *sagitta*, the *asteriscus* and the *lapillus* (Fig. 1.2) . The morphology of the *sagitae* otolith is specially important for the study of taxonomic relationships between species (Nolf 1985, Smale et al. 1995, Tuset et al. 2008). For this reason, a conventional terminology exists to refer to its different parts (Fig. 1.3) and its contour shape (Fig. 1.4)[1]. The sensorial epithelium is joined to the *sulcus acusticus,* which is a cleft that crosses the internal side of the otolith; the *rostrum* corresponds to the top frontal area, the left margin is known as the *caudal margin* (or *postrostrum*); the *dorsal* area is the top part of the otolith; and the *ventral* area is the bottom (Tuset et al. 2008).

More importantly is the fact that otoliths act as natural data loggers, recording information of fish life at different rates related to their growth and environment (Kalish 1991, Campana 1999). This information, which includes age and growth, movement patterns and habitat interactions, can be interpreted in terms of ecology, demography or life history, and has become of fundamental importance in fish management and the protection of species.

[1]Although these definitions cover most of the otolith shapes, otoliths that do not fit exactly to these forms include the word "to" between names to specify that the exact shape is intermediate between the two forms. This notation is assumed in the results section.

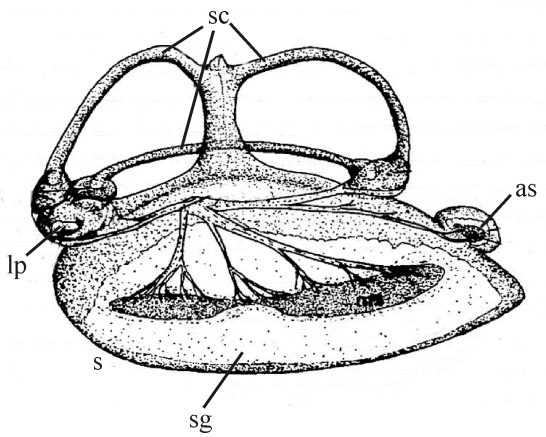


Figure 1.2: Right inner ear of a *Merluccius capensis* specimen. All sensing *maculae* and innervations of the nerves in the epithelium can be observed. as - *asteriscus* otolith; lp - *lapillus* otolith; s, *saccula*; sc - semicircular channels; sg, *sagitta* otolith (from Lombarte 1990).

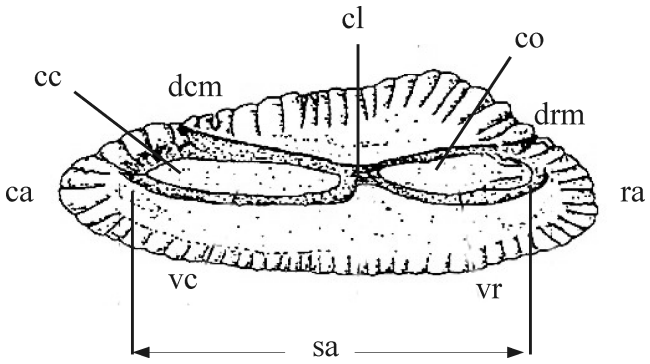


Figure 1.3: Right *sagitta* otolith from a *M. capensis* specimen: ca - *caudal* area ; cc - *collicum caudal*; cl - *collum*; co - *collicum-ostial*; dcm - *dorso-caudal* margin; drm - *dorso-rostral* margin; ra - *rostral* area; sa - *sulcus acusticus*; vc - ventral *caudal*; vr - ventral *rostral* (from Lombarte 1990)

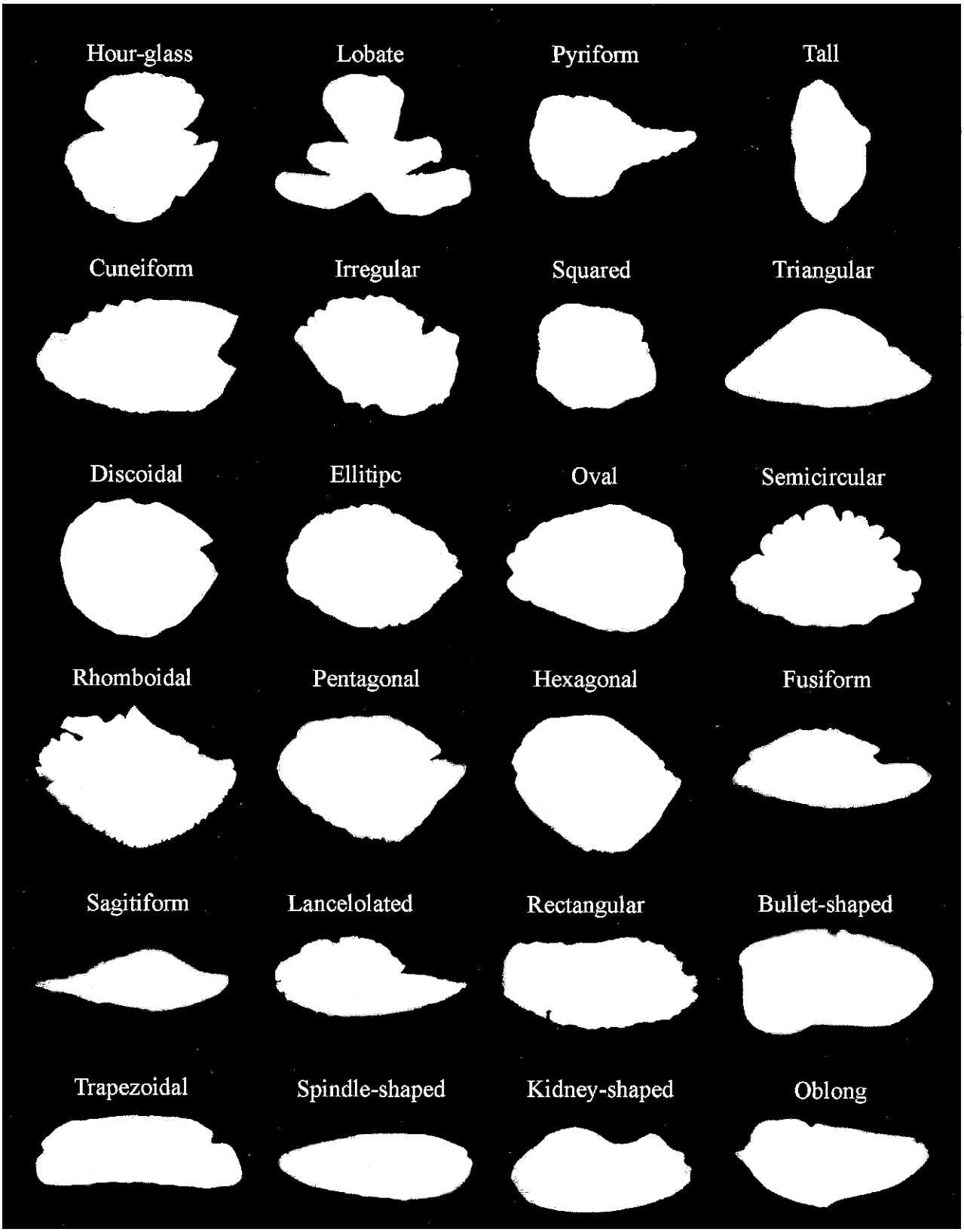


Figure 1.4: Conventional terminology used to describe the shape of otolith outlines (from Tuset et al. 2008).

1.2 The Use of Otoliths in Marine Applications

The field of otolith research and its applications to fisheries science has developed significantly in recent decades, to a great extent because of the technological advances in extracting information (Nolf 1985, Grant 1992, Secor et al. 1995a, Panfili et al. 2002, Campana 2005). Although applications have always been oriented to the sustainability of fish stocks, very frequently they are used to address a broad range of problems, ranging from the management of fisheries to environmental change. These include, for example, the determination of structures within population species (Blacker 1995, Norman and Greenwood 1975, Platt and Popper 1981), inter-specific relationships between different species in biology (Messieh et al. 1989, Lombarte and Castellón 1991, Campana and Casselman 1993, DeVries et al. 2002, Cardinale et al. 2004, Poulet et al. 2004, Castonguay et al. 1991), the identification of species and fossils in paleontology (Slyke 1998, Carpenter et al. 2003), the assessment of population dynamics (Dutil et al. 1999, Finstand 2003), or even the diet of predators (Fitch and Lavenberg 1971, Cottrell et al. 1996, Beneditto et al. 2001).

1.2.1 Identification of Fish Species

The exploitation of marine resources have long depended on inter-specific and intra-specific studies (Wright et al. 2002). In these tasks, the correct identification of species and populations (fish from the same species inhabiting the same area, at a specific time and sharing common morphological and genetic characteristics) is crucial in order to determine and establish not only how they should be managed, but also the fishing policy which should be implemented.

In general, this can sometimes be achieved by using dynamic models, which must be tuned appropriately. Several biological parameters are associated to the models (weight-size relationship, growth, size at maturity and breeding periods, among others) and are used as features in order to identify species correctly. However, in the particular case of *cryptospecies*, fishes which are morphologically similar but genetically different, these parameters are insufficient and additional information from otoliths is necessary to avoid misclassifications that can cause management errors.

Very often, this situation occurs when several species of the same genera share the inhabiting area. For example, in the *Merluccius* genera (common hake), which is considered a taxonomically complicated fish group, the use of otoliths has been proved as highly specific and, therefore, adequate for the identification of the different hake groups in the world (Mombeck 1970, Botha 1971, Lombarte and Castellón 1991, Torres et al. 2000a, Lloris et al.

2005). For this reason, otoliths have acquired an important role in fish identification tasks in recent decades (Tuset et al. 2010).

Although solutions based on taxonomic data are still used nowadays, in general it is accepted that finding an optimal set of otolith shape descriptors from the images, combined with computer analytical tools, provides the best results (Bird et al. 1986, Cardinale et al. 2004, Parisi-Baradad et al. 2010). Nevertheless, the development of automatic tools that allow the characterization of morphological differences may prove more interesting in order to develop more reliable classifications. These characters should, in as far as it is possible, be observable even for the species of the same genera whose irregularities are very cryptic and difficult to verify at a glance (Lloris et al. 2005).

Otolith structures serve to carry out identification tasks, for many reasons: they have specific and population characteristics; they are easy to store and preserve (due to their reduced size); and are available in many fishing institutions and centers for evaluation purposes. On the other hand, a key challenge in the future is how to deal with the great variability existing throughout populations and individual fish growth to design more robust automatic tools.

1.2.2 Aging Applications

Another application of otoliths is related to the translucent and opaque rings that can be observed in otolith sections (or on one side of the entire otolith in some fish species). These can be used to estimate both fish age and growth during its life (Beamish 1979, Nolf 1985, Gauldie 1994, Karlou-Riga 2000, Morita and Matsuishi 2001, Wilson and Neiland 2001, Laidig et al. 2003). The process is very intuitive and consists of counting the potential rings appearing on the surface of the otolith along its main radial (Fig. 1.5). This task is crucial for the management of the seas as more than one million otoliths are read manually every year in order to regulate fishing activities (Campana and Thorrold 2001, Morison et al. 2005).

In a more automated fashion, fish age is best estimated by obtaining a one-dimensional signal from the gray-scale image (Troadec 1991, Campana and Thorrold 2001, Morison et al. 2005). In this sense, computer aging methods address this problem from two different perspectives: *feature-based* and *filter-based* approach.

The first approach is very similar to that of fish identification explained above. After obtaining the main signal of interest from the image (Fig. 1.6) and describing the feature vector of interest, a classification method is used to estimate the age class of the fish. As features from the intensity profile alone are not insufficient to obtain good classification results, other general features from the fish, such as size and weight, are introduced in the classifier. Morison et al. (1998) used a Neural Network to design a system based on reader

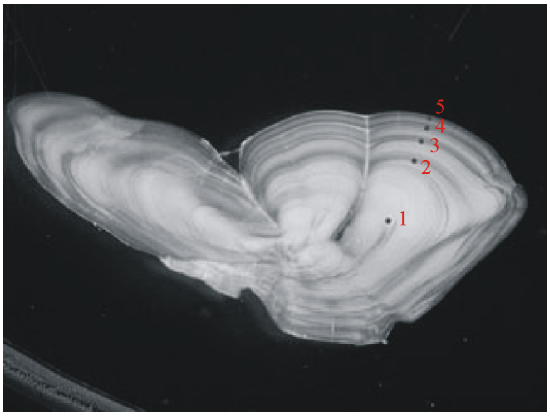


Figure 1.5: Example of manual age estimation for a 5-year cod otolith (source: IBACS European project).

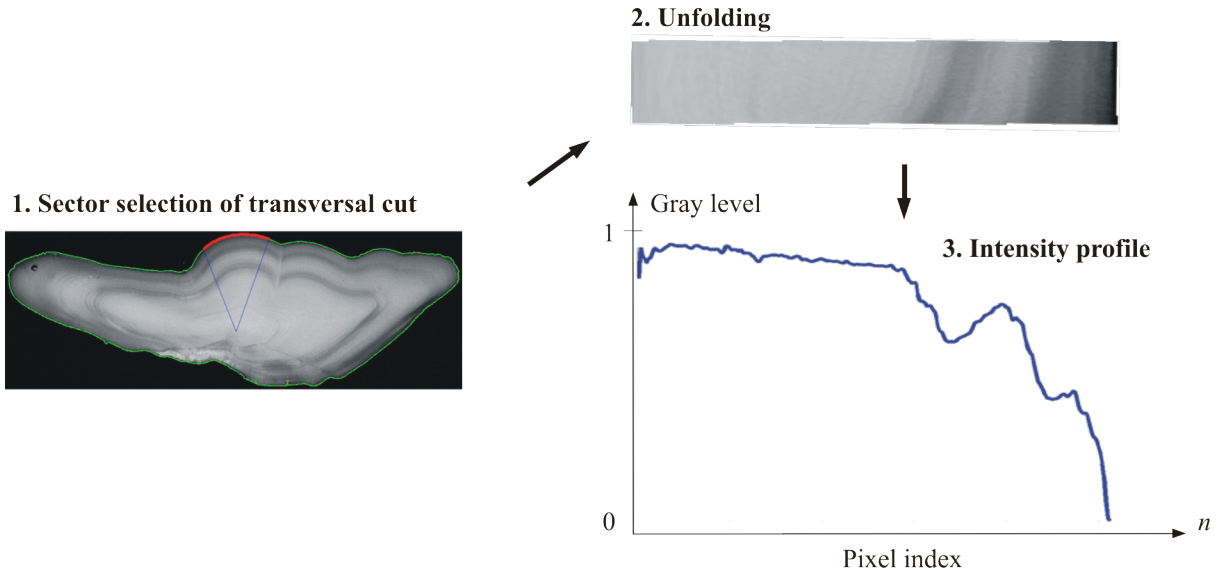


Figure 1.6: Ring extraction example in feature-based methods. Sector selection and unfolding are common processes for both groups of techniques (feature-based and filter-based methods) and require manual implementation. An optimal contrast variation at the otolith border is critical for estimating the age.

experience information and Fourier descriptors. These limitations were demonstrated recently in a study by Fablet and Josse (2005) who stressed the importance of treating age structures beforehand.

The second approach deals with aspects of signal amplification and noise cancellation, among other quality issues of image and signal processing (Lagardère and Troadec 1997, Welleman and Storbeck 1995, Troadec et al. 2000, Guillaud et al. 2002, Fablet 2006). The goal is still to infer the fish age but directly from the intensity profile, after removing the irrelevant structures and compensating growth effects.

In general, aging experiments combine both techniques. In this sense, the most interesting results were obtained by Fablet and Josse (2005) who correctly classified almost 90% (in terms of expert agreement) of plaice otoliths (*Pleuronectes platessa*), ranging from 1 to 6-year old fish. These results were obtained by means of several processing tools, including: image segmentation for canceling contrast; growth compensation techniques based on signal demodulation; peak-based feature description methods combined with the cosinus transform; and a Support-Vector-Machine classifier (SVM). Although these results were satisfactory, solutions for more complex species are aimed at eliminating discontinuity problems arising from the ring profile and the poor contrast by means of statistical learning techniques (Fablet 2006).

The manual estimation of age from fish otoliths has an annual cost of several million euros at a European level, including the training and support from experienced age readers. Despite this expenditure, only about 25% of the ICES (*International Council for the Exploitation of the Sea*) stocks have low uncertainty (Campana and Thorrold 2001, Morison et al. 2005). As such, the development of automatic tools that could improve the objectivity of the reader's interpretation in the estimation of age would prove very useful (Reeves 2003).

1.3 On Pattern Recognition Systems

Taking into account all these problems, it is imperative the design of new standardized methods and software tools that assist in the interpretation of otolith data (Campana 2005). In the field of image analysis and signal processing, the aim of feature extraction is to obtaining short and useful descriptions from the data that can explain the cause of the problem at hand.

This concept, often referred to as the *dimensionality reduction* problem, has intrigued many scientists and several methods have been proposed. These methods, which share the principle of function-cost maximization (or minimization), are generally used to guide the automatic search of features. Feature extraction applications following this principle are

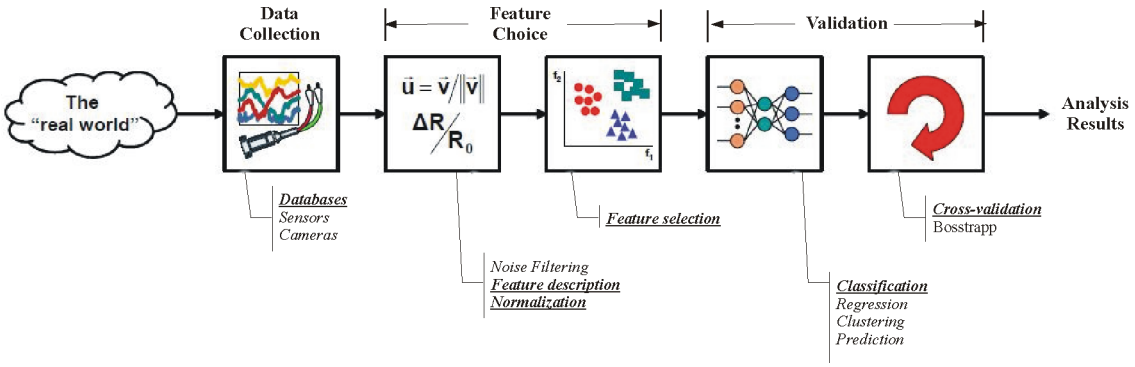


Figure 1.7: Flow diagram of generalized pattern recognition systems (from Osuna 2005). Feature choice and model validation are the main theme of this thesis.

grouped into two main categories:

- **Signal compression:** The goal is to represent the original samples accurately in a lower dimension.
- **Classification:** The goal is to enhance class-discriminatory information in a lower dimension.

In the field of fish identification and aging, clearly, we are interested in the second group. In this case, the process requires different steps which are applied sequentially (Fig. 1.7). Data collection is generally the most time-intensive component of a pattern recognition system and expert knowledge is crucial to determine how to implement the database or which preprocessing scheme is best suited to the application.

Having decided on the structure of data to be analyzed, the next step is to define the descriptive feature models and select those which are relevant for classification. This scenario is formally conceptualized in the following manner: given an input feature space $\boldsymbol{x} \in \mathbb{R}^N$ taken from the original database, determine a mapping $\boldsymbol{y} = \mathcal{F}(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}^K$, where $K<N$, such that the transformed feature vectors $\boldsymbol{y} \in \mathbb{R}^K$ preserves (most of) the *useful* information or structure in $\mathbb{R}^N$.

This procedure relies on a mapping method for representing the features in a different domain and an analytic (or evaluation) strategy to select the most relevant ones. Two different approaches are available for this step: *statistical* (or *decision theoretic* approach) and *structural* (or *syntactic* approach) (Pavlidis 1977, Devijer and Kittler 1982, Fu 1982, Fukunaga 1990, McLachlan 1992).

The main limitation of the statistical approach resides in their coordinate representation system which is very difficult to interpret. There have been two major accomplishments of

this approach: the Karhumen-Loève Transform [also known as Principal Component Analysis (PCA)] - most commonly used in signal compression applications (Jollife 1987); and the Linear Discriminant Analysis (LDA) - which is best suited for classification purposes (Fisher 1936).

Structural methods, on the other hand, are based on the philosophy of "analysis by synthesis". The philosophy underlying this approach is the existence of elementary and intuitive building blocks that constitute the entire signals. For example, the Fourier Transform (FT) describes the signals as a composition of different sinusoids varying in terms of magnitude, frequency and phase components (Cooley and Tukey 1965). Unlike statistical approaches, the functionality of structural methods is purely descriptive, which means that no evaluation of features is developed at all.

Although many description mechanisms base their operation on the estimation of frequency content, very recently the use of methods determining the time-frequency components have gain in popularity, and specially with the *Wavelet transform* (WT) (Mallat and Hwang 1992), which has led to the *Best-Basis* theory (Saito 1994, Saito and Coifman 1995).

Finally, the last stage of the pattern recognition process verifies the selected models by testing their efficiency in the estimation of classes. This operation is supervised by means of classifiers, which must use representative data to obtain consistent results. This does not represent a major problem provided that the number of samples is unlimited but very often data collection is an arduous and expensive process. For this reason, more realistic simulations use part of the data for configuring the classifier (the learning data set) and the remaining samples to estimate its results (the validation data set).

A wide range classification methods exists, and are grouped in two main categories: *Bayesian* (or linear) classifiers which estimate classes based on density functions and, and non-linear classifiers whose performance depends on the structured networks that are used to estimate the class decision boundaries. Examples of the first group are LDA-based (Fisher 1936), Kernel or k-Nearest Neighbor (k-NN, Cover and Hart 1967) classifiers, among others. In the second group, the goal is to construct a network able to detect the class decision boundaries (learning phase) and then use it to infer the classes (validation stage). Examples of this approach are Classification and Regression Trees (CART, Breiman et al. 1984), Neural Networks (NN, Rosemblatt 1957) or Support Vector Machines (SVM, Boser et al. 1992, Vapnik 1995), among others.

Learning Vector Quantization (LVQ, Kohonen et al. 1995), however, is in the middle of both categories. It is also based on a blank network, but in the learning phase it uses the original data to reproduce the true class distributions as faithfully as possible, whereas the validation stage implements a linear classifier (the k-NN approach). As such, it is possible to

combine the advantages of both approaches in one single method.

1.4 Motivations and Scope of the Thesis

From a more physical perspective of the structures of signals, very often we are interested in two main questions: *what* the physical differences in classes are and *how* significant they are. As mentioned above, there is no single method that provides relatively complete answers for both questions. Either we decompose signals in known elementary structures or we evaluate the separability of input data without knowing elementary structures. So the answer to the previous questions often comes from the use of combined approaches.

This scenario has been addressed recently by the *Best-Basis paradigm* (Saito 1994, Saito and Coifman 1995). This feature extraction philosophy consists of three main operations: 1) select an efficient coordinate representation system of input data to solve the problem at hand; 2) sort the coordinates by importance and discard the irrelevant ones; and finally 3) use the surviving coordinates to develop classification tasks. The contention of this work is that such a scenario is more flexible, and better suited towards this goal, than other classical analytical methods only based exclusively on signal compositions of frequency components.

The *Local Discriminant Bases* algorithm (LDB), which belong to this approach, was tested in the comparison of cod and hake otoliths, and for the extraction of age structures, with promising results (Soria et al. 2008). However, its entropy-based criterion had a main drawback. The use of the signal energy information is more suited to compression tasks than those of classification, so the measure may not provide useful information in certain applications. A further refinement improved this behavior by introducing an empirical estimation of feature densities using averaged shifted histograms (ASH) (Cocchi et al. 2001, Saito et al. 2002), and more recently the Earth Mover's Distance algorithm (EMD) has been proposed (Marchand and Saito 2012), but even with these improvements data generalization still remains a problem.

In this sense, it would be preferable to develop discriminant measures of more general behavior. By 'general behavior' we mean measurements specified within boundaries which monotonically increase with the separability of classes, ranging from zero (when features are identical between classes) to one (if they are completely different). This measure should be independent of the description mechanism, as far as it is possible, in order to develop fully automated feature extraction systems.

A new LDB algorithm is proposed in this thesis. This method includes three mechanisms, all aimed at implementing the first two of the feature extraction process: a transformation method for describing time-frequency atoms in signals (i.e. the *discrete wavelet transform*,

DWT, or the *Local cosinus/sinus transform*, LCT/LST), a discriminant analysis based on density distributions, and a feature selection algorithm. Its main advantage is a feature evaluation system of measurements more consistent with the performance of linear classifiers, which makes it possible to select and to predict their classification efficiency before a classifier is used.

A main problem of PCA is that the covariance matrix is more oriented to compression tasks rather than those of classification. A similar inconsistency has been found with the original LDB in the experiments of this work due to the fact that accumulating pairwise measurements among several classes obscures the meaning of discrimination information, and LDA does not have a comprehensive description of the signal components because orthogonality is lost within the description process.

In contrast, our density estimates are based on the use of histograms, which are evaluated by means of normalized and bounded measurements, and a top-down search strategy selects nodes within the space of tree-structured wavelets based on their most representative coefficients.

The main contributions have been restricted to the development of computer tools for otolith-based fish identification experiments. The proposed tools are contrasted to other standard tools used for the same purpose, and also to compare different hake species in real inter-specific and intra-specific experiments. The most significant contributions lie on the interpretation of irregularities: shape, time-frequency position, extension support and degree of importance; which can be highlighted automatically on the same images or signals. For this application, a new preprocessing tool of contour normalization is also proposed in order to reduce the variance translation effects underlying the implementation of the discrete wavelet transform.

Although aging applications are beyond the scope of this thesis, a new approach for compensating the periodicity variation of age structures is presented. This variation is normally caused by several natural phenomena, including environmental conditions of fish habitat and fish genetics, among other factors (Morison et al. 1998, Troadec et al. 2000, Fablet et al. 2003). The developments in this case are based on the Von Bertanlanffy growth modulation function (VBGF) (Beverton and Holt 1957).

Unlike the statistical approaches oriented to finding global fish group parameters, our proposition tries to infer the growth of each single specimen. This strategy is justified because the genetics are quite specific in single individuals, suggesting therefore a more specific method that can cope with such singularities should be used.

The work developed in this thesis regarding this particular aspect is at an early stage. However, it does suggests that to do this, the problem of separating image contrast and

demodulation operations must be dealt with first at the mechanical level of the image acquisition.

Most of the developments included in this document have been supported and funded by the AFORO3D project (Análisis de FORmas de Otolitos 3D, Ministerio de Economía y Competitividad, CTM2010-19701) and the AFISA project from the European Union (Automated FISh Ageing, nº. 044132). In the latter case, some of the software tools and image pattern recognition algorithms developed in this thesis have been integrated in the TNPC software v4.0 from NOESIS (www.noesis.com).

1.5 Overview of this Document

The content has been divided into four main sections. The first section introduces current otolith methodology commonly used in both fish applications, the second one reviews feature extraction methodology in general, the third describes our proposals and, finally, the last section illustrates results and discusses the existing limitations and improvements to be made in future research.

A very general overview of otolith-based fish identification and aging applications, and the pattern recognition steps have been presented in this chapter. Chapter 2 focuses on the technical aspects of otolith preparations, conditioning, image acquisition and data preprocessing. Certain definitions regarding manual aging precision errors will be necessary in order to evaluate the performance of the proposed algorithms, and these definitions are also provided in this chapter.

Chapter 3 reviews different methods for constructing and analyzing feature models. Some of them have been used with some success in otolith fish applications problems, and others are more common in image-audio applications, and their applicability to otoliths has yet to be studied. Of course, the focus is on the Best-Basis paradigm, and in particular the LDB, which is the object tool of our proposals, described in chapter 5.

The developed tool is called *Density Local Discriminant Bases* (DLDB) because of the new way to represent the separability of class distributions. As this feature extraction mechanism is expected to be equipped with better generalization capabilities than previous proposals, we give a more formal definition of what 'generalization' means. The methods for the otolith contour normalization and the demodulation of intensity profiles are also presented in chapter 5.

Chapter 4 provides the basics of classifiers, necessary in the validation of feature models and discriminant measures. Former definitions for 'supervised learning' and the validation stage are provided. Although not all the classification methods can be covered, because there

are too many, comparisons between linear and non-linear methodologies are made.

Experimental results are provided in chapter 6. The first experiment, shows the advantages of the DLDB compared to standard tools used in the field of otolith-based fish identification. In particular, its ability for detecting and explaining signal irregularities and its consistency with identification results are proved. A further experiment uses DLDB to compare the morphological otolith differences of *Merluccius* species. Conclusions about the biological differences between hake species are provided. The chapter finishes with the results on aging. The demodulation ability of our preprocessing tool is demonstrated and age estimation results for plaice otoliths (*Pleuronectes platessa*) are discussed.

Of course, there are always limitations to every methodology and the proposals of this work are no exception. The last chapter outlines some issues to be addressed in the future. A possible improvement of DLDB is proposed and formalized, but the software still has to be implemented.

Chapter 2

Acquisition of Otolith Data

2.1 Introduction

This chapter deals with the more technical operations involving the first and second stage (see Fig. 1.7), and corresponds to the preparation of otoliths, their mechanical operations, image acquisition techniques, segmentation algorithms and preprocessing techniques. All them are intended to generate useful otolith digital data for its analysis.

While these operations are standard procedures well accepted today in this field, in the case of aging our proposals will be purely experimental (see chapter 5). In general, the identification of fish species lies in the use of otolith contour data, which is much easier to obtain than the gray-scale data of aging applications.

In fish identification, an image with a nearly constant background is sufficient to extract contour points in favorable conditions, while in aging applications the chemicals used for the preparation of otoliths, light conditions and the configuration of the camera, which conditionate the image quality, play an important role in the calculation of intensity profiles.

These issues are considered in section 2.2. In addition to the factors already mentioned, the discussion includes considerations on the conservation of otoliths and the manual validation of age readings. The latter is necessary to evaluate the performance quality of automatic aging tools. Sections 2.3 and 2.4 are more concerned with the digitization of contour data points and gray-scale one-dimensional intensity profiles.

For the extraction of the contour points, any standard segmentation method may be valid since light conditions are not too restrictive. Contour information also serve to delimit the enclosing area age structures will be extracted from. Two preprocessing operations are necessary to extract the intensity profile: 1) the detection of the otolith nucleus and, 2) the preprocessing of translucent and opaque rings to obtain the one-dimensional signal. The first

issue has been addressed successfully by Cao and Fablet (2006) with an algorithm that has proved to be quite robust. The second issue, however, is currently under research. The work developed by Fablet and Josse (2005) for plaice otoliths constitutes the main starting point of our proposals.

Finally, section 2.5 summarizes the relevant points and looks at the present limitations and future challenges of current experimental work.

2.2 Basic Otolith Preparation

2.2.1 Direct Observation

While otoliths can be observed directly in their original state with a simple optical magnification device, the development of analytical tasks based on computer-assisted imaging techniques require certain preparations.

In general, otoliths are typically observed directly or immersed in a liquid observation medium, configuring proper lighting conditions, and using a low-power binocular microscope with an embedded camera. When the full otolith is used, it is immersed in water, alcohol or oil; and observed with either transmitted-light or reflected-light techniques. The oil helps to highlight the structures from the otolith surface and avoids reflected glare. Historically, these oils have included clove oil, "baby oil" or creosete, though the latter is no longer used. It is also important to use a cold light source to prevent the evaporation of oil during the observation.

This procedure is suffice even for counting the rings from *sagittae* of flatfish and certain pelagic species, since their otoliths do not require sectioning.

2.2.2 Embedding for Incident-light Observations

Despite the simplicity of direct observation, otoliths are also elements included in a wide range of databases and must be properly stored to prevent them from deteriorating, and to ensure that any sectioning, grinding and/or polishing can be carried out safely at a later date.

The operations necessary for their conservation form part of a process commonly referred to as 'embedding'. This process can be complex and expensive, and requires of different laboratory tools and materials which need to be maintained in good working order. For example, the resins required to immerse the otoliths, become opaque as they deteriorate and so must be stored in cool dark places. Table 2.1 summarizes the principal embedding media

Material	Uses	Advantages	Disadvantages
Polyester resin (PR)	- high-quality permanent fixing - embedding and surface grinding	- multi-purpose -moderate optical clarity -good grinding/ polishing properties - light stable	- non-reversible - variable setting quality - slow setting (12-24h) - toxic - low shrinkage - catalyst required
Epoxy resin	- high-quality permanent fixing - embedding and surface grinding	- harder than PR - high optical quality - light stable - low chemical contamination - minimal shrinkage - wide range of applications	- non-reversible - moderate/slow setting toxic expensive catalyst required
Thermoplastics	- high-quality - permanent fixing - embedding and surface grinding	- fast setting (minutes) - easily removed - high optical clarity - non-toxic - good grinding/ polishing properties	- heat needed (70-160^{o}C) - develops gas bubbles - moderate shrinkage - high Si content
Super Glue (cyanoacetate)	- rapid setting high strength fixing - surface grinding	- easy to use - sets without heat or chemical hardeners	- high S, Pb content - can detach from slide
UV dental glues	- rapid setting high strength fixing - surface grinding	- sets without heat or chemical hardeners - non-toxic	- very expensive - optical clarity not known
Wax	- quick temporary fixing - embedding only	- fast - cheap - easily removed	- soft - poor optical properties - non-permanent
Eukitt	- good quality soft fixing medium - embedding only	- easy to use - no hardener - good optical clarity - easily removed	- slow setting - remains soft - unstable over time - can detach from slide

Table 2.1: Standard mounting materials (from Mosegaard et al. 1998)

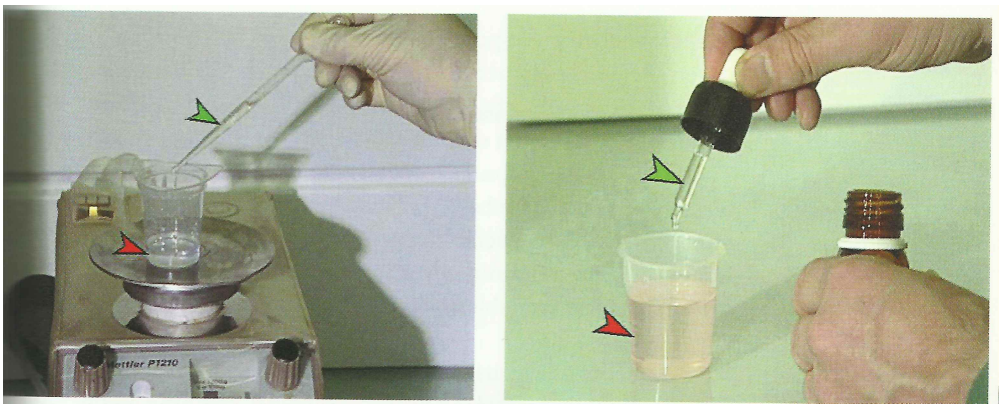


Figure 2.1: Preparation of the catalyst. The volume required depends on the full amount of polyester resin required according to Table 2.2 (from McCurdy et al. 2002)

Resin volume (ml)	**Resin** weight (g)	**Number of drops of catalyst** 1% mixture	2% mixture
91.0	∼100	20	40
45.5	∼50	10	20
23.3	∼25	5	10
11.5	∼12.5	2.5	5
9.1	∼10	2	4

Table 2.2: Simple dosage of polyester resin (in volume and/or weight) and catalyst (in drops) depending on their ratio. The greater the quantity of catalyst the higher the speed of setting (from Mosegaard et al. 1998)

and their applicability. The most frequently used resins are the hard and irreversible synthetic ones (polyester, epoxy, etc). In general, the embedding operation proceeds as follows:

The first operation consists in the polymerization of the polyester by mixing the polyester with 1-2% of catalyst (i.e. 25g of polyester resin and 0.2g of catalyst, Fig. 2.1). Table 2.2 provides some information on the proportions of quantities to be mixed, depending on this ratio. Traditionally, the resin was dissolved in syrene, and the catalyst used an accelerator to reduce the hardening time. However, as there is risk of explosion, current resins are supplied pre-accelerated.

All the media with the catalyst should be mixed carefully in order to prevent bubbles forming in the mixture, as they complicate the observation of calcified structures. The mixture is then left to rest for a period of time to allow the biggest bubbles to burst. The hardening time varies from a few seconds (Super Glue or UV dental glue) to a few hours (epoxy or polyester) depending on the media used. A dry oven is often used to complete the

polymerization process and ensure the mixture has turned into gel. The polymerization of polyester requires at least 24h in a dry oven (around 30^{o}C) for optimum results.

Resins can be colored by adding a pigment, generally black, or they can be purchased pre-colored. When otoliths are embedded for posterior observation a high-transparency polyester resin is normally used. Resins and catalysts are often known as 'hardeners' and are available, along with mounting kits, from specialist suppliers.

When the polymerization is completed, the next step is to place the calcified structures in cylindrical cavities on black plastic slides. Black slides provide better contrast for observations of *annuli* in the case of age estimation. The slides are 65mm x 60mm in size, cut from 3mm-thick Perspex^{TM} sheet, and are normally machined in grids of 25 flat-bottomed cavities, each 7mm in diameter and approximately 1.2mm deep. This provides sufficient room for a reference number to be engraved on each slide with a hardened steel stylus or an electric engraving tool.

Entire otoliths are positioned in the cavities on the plastic slide in pairs, with the *sulcus acusticus* facing downwards and, therefore, not visible (Fig. 2.2a). Occasionally, small otoliths acquire static electrical charges, making it more difficult to position them in the cavities. These problems can be resolved by increasing the humidity of the atmosphere or changing the type of clothing worn during the placement operation.

A plastic pipette is used to drop the catalyzed clear resin into the cavities, covering the otoliths and ensuring that the floating ones are not re-positioned (Fig. 2.2b). The catalyzed resin is dropped between otolith pairs to form a single layer of resin that completely covers all otoliths on the slide. A custom-sized glass microscope cover slip is then floated over the resin (Fig. 2.2c) and left in the fume cabinet until completely set (Fig. 2.2d). Cover slips generally deform above the cavities and may crack as the thicker layer of resin contracts further on hardening. The resin will set firmly within a few hours, but observations must not begin until the resin has set and there is no risk of styrene being absorbed through the skin.

Perspex^{TM} is expensive and a number of alternative plastics that bond equally well with hardening resin are currently used, and injection-molded slides with embossed numbers are commercially available at cheaper prices. When using machined slides, the plastic must be hard enough and the drill sufficiently sharp to ensure the base of each pit is smooth and level. If this is not done, thin patches of partial vacuum may form between the hardened resin making it difficult to observe calcified structures,since they will have the silvered effect of a mirror.

These operations ensure that the otolith can be stored safely and used in future observations and is also valid for very small calcified structures, including many flatfish species (except the thicker otoliths) whose ring structures can be observed without the need of sec-

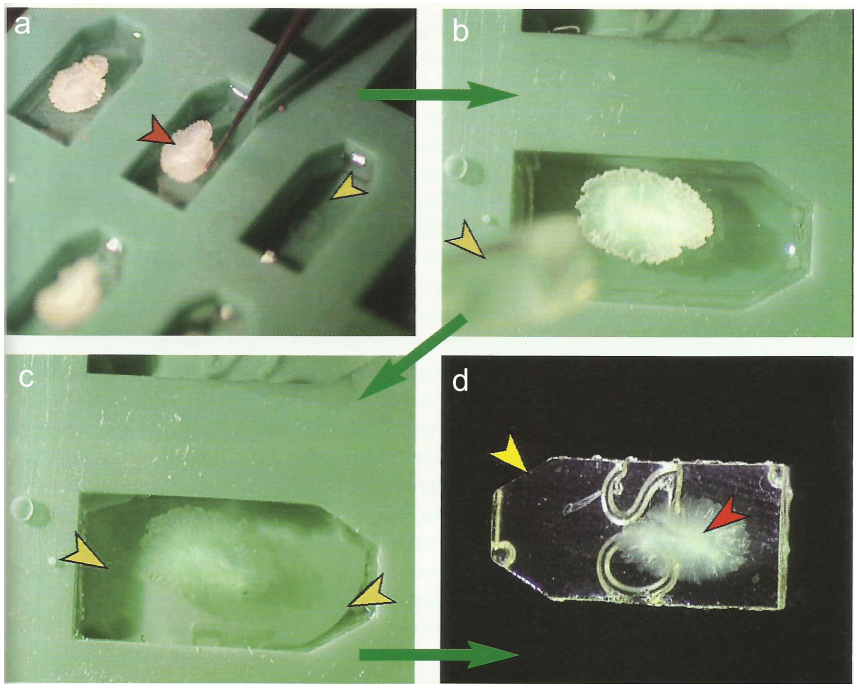


Figure 2.2: Example of otolith embedding in an elastomer mould. a) A polymerized resin layer (yellow arrow) is run on the bottom of each location before the otolith is deposited on the bottom. b) The otolith is oriented and embedded in a second layer. c) It should be made sure that the resin fills the mould completely. In this case the resin does not reach the edge so that it can be turned over to drive out the air bubbles. d) Otolith definitively embedded (red arrow) (from McCurdy et al. 2002)

tioning operations (Parrish and Sharman 1959, Watson 1965). In such case, the operations and materials differ considerably from those presented in this section. The reader is referred to McCurdy et al. (2002) for the details.

2.2.3 Manual Validation of Age Readings

Since age information forms the basis of most influential biological variables (growth rates, fish stocks, mortality and also fisheries productivity, among others) a variety of methods exists through which manual age interpretations can be validated. These methods can be classified as either validating absolute age, validating the periodicity of growth increments or of corroborating an existing set of age estimates.

Generally speaking, two types of errors can occur when fish age is manually estimated: 1) errors that affect *accuracy* of age readings, or closeness to the age estimate of the true value, and 2) errors that affect the *precision* or the reproducibility of repeated measurements on a

given structure by the reader.

The first type is more associated to the limitations of the processes and techniques used in the detection of ring structures, given that not all of them belong to the growth sequence of the fish. Thus, it is a deterministic error that the expert can reduce, or even eliminate, by using the most appropriate validation method in each situation. In general, their goal is to validate age structures, either by the assessment of the true absolute age, the periodicity of growth increments.

In this sense, *known-age-release* (Tuskamoto and Kajihara 1987, Quinn et al. 1991, Campana 1999, Secor et al. 1995b), *bomb radiocarbon* (Kalish 1993, 1995a,b, Campana and Jones 1998) or *radio-chemical dating* (Bennett et al. 1982, Campana et al. 1990, Burton et al. 1999) are very used in the first case; whereas *mark-recapture of chemically-tagged* fish (Geffen 1982, Oliveira 1996, Natanson et al. 2002), *discrete length modes* (Morales-Nin 1989, Hanchet and Uozumi 1996, Natanson et al. 2002), *natural* and *data-specific markers* (Blacker 1995, Campana and Neilson 1985, Wilson and McCormick 1997) and *marginal increment analysis* (MIA) (Hyndres et al. 1992, Morales-Nin et al. 1998a, Carlson et al. 1999) are well known in the second case.

Very briefly, the idea of known-age-release is to know the absolute age of recaptured fish by calculating the incremental difference in relation to the time of release. This technique is very used in juvenile specimens, as the exact age is determined without error whereas bomb radiocarbon, although very expensive, is preferred for the older specimens because of their lower recapture rates. Alternatively, mark-recapture of chemically-tagged fish is best for validating the periodicity of growth increments. The method is based on rapid incorporation of calcium-binding chemicals, which are applied at the time of tagging into bones, scales and otoliths. The result is a permanent and visible mark under fluorescent light at the time of tagging which serve as a reference for determining posterior daily and annual growth increments (Natanson et al. 2002). The reader is referred to Campana (2001) for an extended summary of manual age validation techniques.

Precision errors, on the other hand, are associated with the reader subjectivity at the time of counting and interpreting growth increments, and they can be biased or random depending on whether the factors associated to the reader occur frequently or sporadically. A common way to estimate this error is that several readers develop repeated age estimations of the same specimen.

The average percent error (APE) has been the traditional index for this measure. However, since many authors pointed out its inadequacy (Beamish and Fournier 1981, Chang 1982, Campana et al. 1979), and subsequently Beamish and Fournier (1981) introduced a modification, defined as

$$APE = \frac{100}{n} \times \sum_{j=1}^{n} \left[\frac{1}{r} \sum_{i=1}^{r} \left(\frac{|t_{ij} - t_j|}{t_j} \right) \right], \tag{2.1}$$

where t_{ij} corresponds to the i-th age determination of the j-th fish for a total of n fishes, t_j is the mean age estimate of the j-th fish, and r is the number of times each fish is read by different readers. Chang (1982) suggested substituting the absolute deviation by the standard deviation from the mean age. So the resulting equation produces an estimate of the coefficient variation (CV), expressed as:

$$CV_j = \frac{\sqrt{\sum_{i=1}^{r} \frac{(t_{ij} - t_j)^2}{r-1}}}{t_j}, \tag{2.2}$$

where CV_j is the age precision estimate for the j-th fish. The index of precision D is similar to CV but calculated as

$$D = \frac{100}{n} \times \sum_{j=1}^{n} \frac{CV_j}{\sqrt{r}}. \tag{2.3}$$

In general, the design process of computer-based aging tools are mainly concerned with accuracy errors. As in the manual case, results have to be provided in terms of the reader's precision, since supervised classifiers also require knowing class information to be properly configured (Campana 2001).

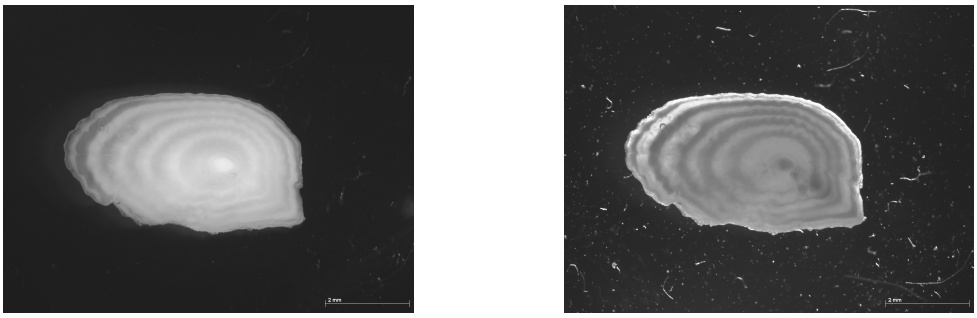
For example, assume that the accuracy of classification experiment is 88% in relation to an inter-expert agreement of 95%. The expected success will be 83.6% taking into account both figures as part of the process of age estimation.

2.2.4 Image Quality Issues

Although images must meet a minimum quality for the relevant structures to be represented optimally with digital data, their memory size do not need to large, since the control of environmental under an homogeneous black background allow gray-scale colour images to be used.

In practice, a minimum resolution of 640x480 pixels using gray-scale suffices for both tasks. The specification of contrast, however, acquires a critical role in automatic aging, as this parameter needs to be optimally configured in several ways for different images not only to highlight potential rings belonging to fish age but also to ensure that the irrelevant rings are removed.

Images can be obtained by using either *transmitted* or *reflected* light (Fig. 2.3). In the



(a) (b)

Figure 2.3: The same plaice otolith image captured for aging tasks. (a) *reflected-light* and (b) *transmitted-light*, respectively. Gray-scale levels corresponding to translucent and opaque regions are inverted.

transmitted mode, the source and the camera lie on opposite sides of the otolith. The light waves corresponding to the opaque regions pass through the bulk and are re-emitted to the camera. In reflected mode, however, opaque and translucent regions are inverted, since light source and camera share the same side, and the lenses catches the beams corresponding to the brighter zones reflected by the otolith (Lombarte et al. 2006).

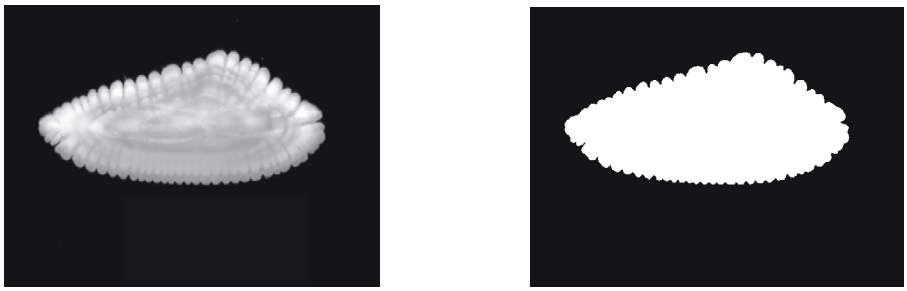
The transmitted mode is normally preferred in aging applications because it offers a higher capacity of contrast variation between translucent and opaque zones, which allow age structures to be better characterized by means of computer programs, but its major drawback is that it requires more complex and expensive preparations compared to reflected light. For this reason, it would be very interesting the development of an efficient aging system based on the reflected mode.

2.3 Contouring by Segmentation

Identification applications base their success on the analysis of contour coordinate points, which are also digitized after removing all internal structures.

This operation is performed by changing the gray-scale image into a binary representation, thanks to the high contrast found at the edge of the otolith. Here, we segment two regions by identifying the pixels that share the properties of both regions. A natural way of doing this comes with the intensity level, which is the simplest property of a pixel. Then, we can use a threshold to separate light and dark regions.

Thresholding is considered one form of segmentation and is a common task in image processing. It creates binary images by turning all pixels below some intensity value to zero



(a) (b)

Figure 2.4: Example of image segmentation using the Otsu method: (a) Original gray-scale image; (b) Binarized image.

and all pixels above that threshold to one (Fig. 2.4). That is,

$$g(x) = \begin{cases} 1, & \text{if } u(x) \geq \lambda_t \\ 0, & \text{otherwise} \end{cases}, \tag{2.4}$$

where $g(x)$ is the thresholded version of the gray-scale image u(x) at intensity level λ_t, for $\{\lambda_t; \lambda_0 = 0, \ldots, \lambda_T = 1\}$ and $x \in \mathbb{R}^2$.

The majority of methods develop an automatic selection of λ_t. One possibility is to set the threshold by minimizing the *within-class* variance which is defined as,

$$\sigma^2_{\text{Within}}(\lambda_t) = n_B(\lambda_t)\,\sigma^2_B(\lambda_t) + n_O(\lambda_t)\,\sigma^2_O(\lambda_t) \tag{2.5}$$

where $n_B(\lambda_t)$ and $n_O(\lambda_t)$ are cumulative histograms of the pixels in background and foreground, respectively. That is, $n_B(\lambda_t) = \sum_{i=0}^{t-1} p(i)$ and $n_O(\lambda_t) = \sum_{i=t}^{T} p(i)$. Bear in mind that a histogram is a probability distribution of the form $p(i) = n_i/N$, where n_i is the number of pixels having intensity λ_i and N is the total number of pixels.

Computing the within-class variances for both classes and for each possible threshold involves a great deal of computation. Otsu showed that minimizing the intra-class variance is the same as maximizing inter-class (or between) class variances (Otsu 1979) and thus, the optimal threshold is obtained as

$$t* = \arg\max_t \left\{\sigma^2_{\text{Between}}(\lambda_t)\right\}, \tag{2.6}$$

where

$$\sigma^2_{\text{Between}}(\lambda_t) = \left\{n_B(\lambda_t)\,n_O(\lambda_t)\left[\mu_B(\lambda_t) - \mu_O(\lambda_t)\right]^2\right\}. \tag{2.7}$$

Here,$\mu_B(\lambda_t)$ and $\mu_O(\lambda_t)$ are the cluster means. So, for each potential threshold, λ_t, we separate the pixels into two clusters, calculate their means, square the differences between means and multiply by the number of pixels in one cluster times the number in the other. Note that the calculations are dependent as we change from one threshold to another, since all the quantities can be updated by simple recurrence relations as pixels move from one cluster to another. That is,

$$n_B\left(\lambda_{t+1}\right) = n_B\left(\lambda_t\right) + n_t, \qquad n_O(\lambda_{t+1}) = n_O\left(\lambda_t\right) - n_t \tag{2.8}$$

$$\mu_B(\lambda_{t+1}) = \frac{\mu_B\left(\lambda_t\right) n_B\left(\lambda_t\right) + n_t\lambda_t}{n_B\left(\lambda_{t+1}\right)}, \qquad \mu_O\left(\lambda_{t+1}\right) = \frac{\mu_O\left(\lambda_t\right) n_O\left(\lambda_t\right) - n_t\lambda_t}{n_O\left(\lambda_{t+1}\right)}, \tag{2.9}$$

where n_t is the proportion of pixels with intensity level T.

After image binarization, the contour silhouette is characterized by the coordinate points at precisely the border of both regions. These points must change into a relative reference axis, so the otolith center is estimated averaging the values in both coordinate axis. Let $\{x\left(n\right), y\left(n\right)\}$ denote coordinate points for $n \in (0, \ldots, N-1)$, that is $\mu_x = \frac{1}{N}\sum_{n=0}^{N-1} x\left(n\right)$ and $\mu_y = \frac{1}{N}\sum_{n=0}^{N-1} y\left(n\right)$. Then, the otolith radials are calculated as,

$$\rho\left(n\right) = \sqrt{\left(x\left(n\right) - \mu_x\right)^2 + \left(y\left(n\right) - \mu_y\right)^2}. \tag{2.10}$$

In practice, when different otolith pictures are processed with this framework, data will not share the same properties of size, image orientation and starting point of contour samples. This situation arises because such conditions are very difficult to control mechanically during the acquisition of images. As such, a preprocessing method of contour standardization, prior to feature extraction, will be necessary.

2.4 Extracting Age Signals

It was pointed out in subsection 2.2.4 that the environmental conditions in which otoliths are photographed play an important role in the detection of age. Light reflection, otolith position or alignment may cause some parts to be brighter (in the light) and opaque (in shadow) in ways that have nothing to do with the real objects in the image and, therefore, may produce signals with useless structures.

When this situation is controlled and the ring structures are well captured in the image, the extraction of intensity profiles comprises three main steps: 1) the determination of the

otolith center from the image (also known as the nucleus); 2) the calculation of the gray-scale values along the radials; and 3) the preprocessing of growth information in order to demodulate the periodic age structures of this signal (Morales-Nin et al. 1998b).

2.4.1 Nucleus Detection

The automatic detection of the nucleus was first considered in Wellman and Storbeck (1995), where looking for the darkest point centered on a predefined area of the image was proposed. This simple approach was not as robust as expected and, later, Guillaud et al. (2002) developed a detection method based on the use of agent systems. Here, the center was calculated as a byproduct of the extraction of two-dimensional rings. Although interesting, the configuration of agents involved a very complex parametrization for it to work appropriately.

Recently, this issue has received thorough attention from Cao and Fablet (2006) in a method that has proved very robust for transmitted light mode. This approach addresses nucleus detection from the perspective of topographic maps, which provide an appropriate framework for describing otolith structures at different gray levels (Monasse and Guichard 2000).

Let $\boldsymbol{u}$ be a gray level image. An upper gray level, set at value l, is characterized by the set of pixels,

$$\mathcal{X}_\lambda(\boldsymbol{u}) = \left\{x \in \mathbb{R}^2, u(x) \geq l\right\}. \tag{2.11}$$

A topographic map is then defined as the collection of all image sets of different levels up to an increasing contrast change. Each level contains the interior part of all connected points that form a shape, and the superior shapes are those that come from higher levels (Caselles et al. 1999). With this definition in mind, the nucleus is then detected as an *elliptical* shape of a *dark* region, *not too eccentric*, *close* to the semi-major axis of the otolith. Fig. 2.5 illustrates the main steps that select the optimal nucleus. This is a two-pass flow which eliminates structures sequentially until the final nucleus candidate is estimated.

2.4.1.1 Maxima Removal and Grain Filtering

Let S_0 be the otolith itself extracted as the largest shape in the image and let $\mathcal{E}_0$ be the set containing all the shapes included in S_0. Also, let $\mathcal{E}_1$ denote the set of potential candidate shapes for the nucleus in $\mathcal{E}_0$. These exclude all the bright shapes and the localized dark shapes of a very small area. Of course, this is rather strict, since spurious noise creates many of them.

Therefore, a grain filter (Vincent 1993, Monasse and Guichard 2000) is used to ensure that the dark region enclosing the nucleus is not eliminated, before the elimination of bright

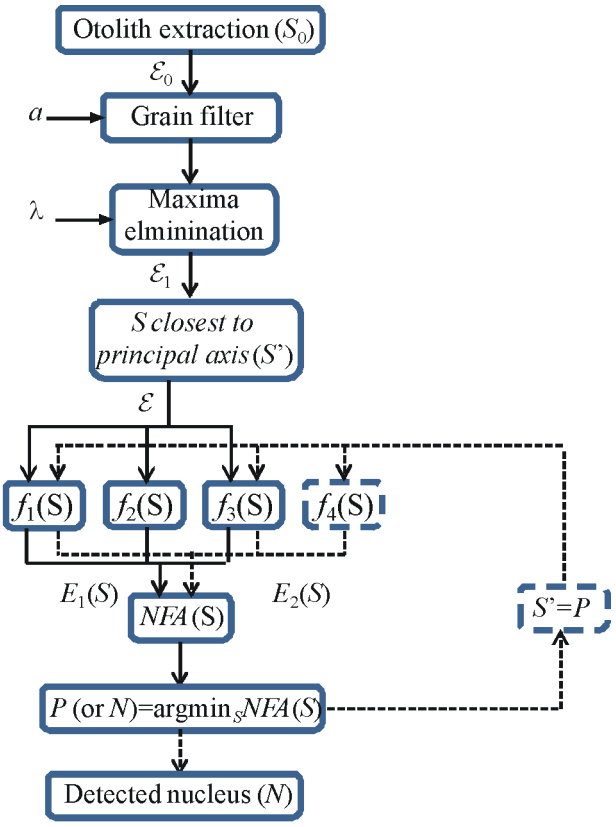


Figure 2.5: Nucleus detection algorithm scheme (from Cao and Fablet 2006). A dotted line is used to indicate the 2nd pass by which the empirical probability $E_2(S)$ is calculated and the final shape for the nucleus (N) determined.

shapes (maxima elimination). The threshold parameter, a, specifies the size of reference (in pixels) which is used to remove the small structures. A few dozen dark shapes remain as possible candidates for the nucleus after these two filtering operations.

2.4.1.2 Principal Axis Proximity

While maximal removal eliminates the bright shapes this step removes the elongated dark shapes enclosing the nucleus, or of long distance to the principal axis, which definitely are opposite to the morphological description provided above (see Fig. 2.6) .

Let $\triangle$ denote the first principal otolith axis. Also, let D be a perpendicular axis to $\triangle$, and $\{X_S(i)\}_{i\in I}$ be a finite family of connected points along D intersecting in shape S ($D \cap X_S(i) \neq \emptyset$). Then, the closest shape to the semi-major axis, S', is determined as

$$S' = \arg\min_{S\in\mathcal{E}_1} d\left(D\cap\triangle, \min_i \{D\cap X_S(i)\}\right), \tag{2.12}$$

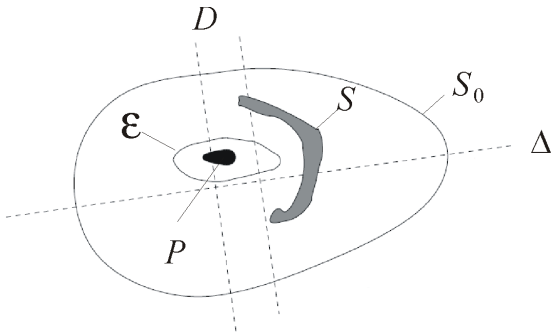


Figure 2.6: Proximity to principal axis (from Cao and Fablet 2006). The black-filled shape is in $\mathcal{E}$ since in any direction orthogonal to $\triangle$, it is contained in the shape closest to $\triangle$ (the white one). For the same reason, the white shape belongs to $\mathcal{E}$. On the contrary, one can find a normal to $\triangle$ such that the gray-filled shape is not closest to $\triangle$ (since the white is closer). Therefore, it does not belong to $\mathcal{E}$.

where $d(A,B)$ denotes Euclidean distance between points A and B. The connected points belonging to this shape will then establish the boundaries that will enclose potential candidate shapes for the nucleus .Let $\mathcal{E}$ be the set containing all the shapes included within these boundaries.

Definition. 2.1 $S \in \mathcal{E}$ if $S \in \mathcal{E}_1$ *and if for any axis* D, *perpendicular to* $\triangle$ *such that* $D \cap X_S(i) \neq \emptyset$, *then there exists* $S' \in \mathcal{E}_1$ *such that S' has minimal distance to* $\triangle$ *on* D *among the shapes of* $\mathcal{E}_1$ *and* $S \subset S'$.

The interpretation of this definition is very simple. Basically, if we start from $\triangle$ and move along its normal, then the first dark shape we encounter must contain potential candidates for the nucleus.

2.4.1.3 Geometrical Statistical Selection

The last operation consists of selecting the most probable candidate shape for the nucleus. For this purpose a geometrical evaluation of the remaining shapes is developed. These geometrical properties are represented by means of three functions $(f_i; i = 1, \ldots 3)$. They are 1) the *area*, defined as

$$f_1(S) = |S|, \tag{2.13}$$

2) the *proximity* of S to $\triangle$ which is measured relatively to the shape size

$$f_2(S) = \frac{|S|}{\max_i d(X_S(i), D \cap \triangle)^2}, \tag{2.14}$$

and, finally, 3) the *flatness*

$$f_3(S) = \frac{|S|}{L(S, \triangle^{\perp})^2}, \tag{2.15}$$

where $L\left(S, \triangle^{\perp}\right)$ corresponds to the diameter of the projection of S on the orthogonal direction of $\triangle$.

According to Cao and Fablet (2006) a good nucleus should yield large values of f_i. However, in order to avoid the use of thresholds for the f_i, which may make the algorithms numerically too sensitive, instead they develop *a contrario* hypothesis that the f_i are statistically independent. This idea is based upon the fact that size, proximity and flatness have a common cause in any shape except for the nucleus, since generally otoliths grow outwards in any direction from the nucleus.

In statistical theory, this is equivalent to testing the hypothesis $\mathcal{H}_0$, in the event that "the f_i are independent", and then making decisions according to the 'null $\mathcal{H}_0$' assumption (otherwise, verifying $\mathcal{H}_1$, "the f_i are not independent", would require a good otolith model which is not easy to compute). A natural way to detect this violation of independence is to use the shape values of the previous geometrical functions as thresholds to determine the expectation of the event $f_i \geq f_i(S)$. Let F_i denote this value which is represented by the empirical inverse repartition function of f_i:

$$F_i(\lambda) = \frac{\#\{f_i \geq \lambda\}}{\#\{\mathcal{E}\}}, \tag{2.16}$$

where $i \in [1, 3)$ and $\#$ denotes cardinality. Also, let

$$E_1(S) = \left(\max_i F_i(f_i(S))\right)^3 \tag{2.17}$$

define the empirical probability to observe the event $f_i \geq f_i(S)$ under the *a contrario* independence hypothesis, computed for each shape $S \in \mathcal{E}$. The best nucleus candidate will then be obtained as the least probable under this independence assumption, but we will need to be sure that this probability is significantly small. One possible way to address this matter consists of calculating the 'number of false alarms' (NFA) of each S, which is defined as

$$NFA(S) = \#\{\mathcal{E}\} E_1(S). \tag{2.18}$$

Thus, a shape is said to be $\varepsilon - meaningful$ if $NFA(S) < \varepsilon$, where ε is an error threshold representing the expected number of shapes permitted outside the a contrario hypothesis. Indeed, Cao and Fablet (2006) proved that the expectation of the number of $\varepsilon - meaningful$

shapes in the independence model is always smaller than ε. As a consequence, a good strategy is to set $\varepsilon < 1$ and check the condition $NFA(N) < \varepsilon$.

Now, we are able to provide a formal definition for the nucleus:

Definition. 2.2 *Let S be a candidate shape among all candidates in $\mathcal{E}$, $S \in \mathcal{E}$ for the nucleus P. This shape is said to be $\varepsilon-$meaningful if, and only if,*

$$P = \arg\min_{S \in \mathcal{E}} NFA(S), \tag{2.19}$$

provided that the condition $NFA(S) < \varepsilon$ holds, for $\varepsilon < 1$.

In practice, however, there will be more shapes that meet these specifications of nucleus, so an additional step is developed using the detected shape in (2.19) to refine the previous prediction P of the nucleus. This refinement assumes that the otolith grows asymmetrically and the distribution of the nucleus position obtained from all selected shapes is clearly unimodal.

Therefore, an additional feature is introduced as a second pass of the *a contrario* framework to select a more precise nucleus. This feature evaluates the distance between the center of the candidate region and that of the predicted nucleus in the first pass

$$f_4(S) = -\|C(S) - C(P)\| \tag{2.20}$$

where $C(\cdot)$ denote center position calculated as the median value of the coordinate points of S and P, respectively. However, since the horizontal orientation of the otolith is unknown the preferred calculation is given by

$$f_4(S) = -\min\left(\|C(S_L) - C(P_L)\|, \|C(S_R) - C(P_R)\|\right) \tag{2.21}$$

where the index, L and R, denote position towards the left and right frame of the principal axis, respectively. Thus, in this second pass the empirical probability of the *a contrario* hypothesis of independence is calculated as

$$E_2(S) = \left(\max_{i=1,3,4} F_i(f_i(S))\right)^3 \tag{2.22}$$

and used in 2.18 and 2.19 to determine the nucleus, N. Note that the independence criterion is not considered for all $i \in [1,4)$, given that f_2 and f_4 are obviously dependent. On the other hand, the independence between f_1, f_3 and f_4 is more reasonable in the *a contrario* detection approach.

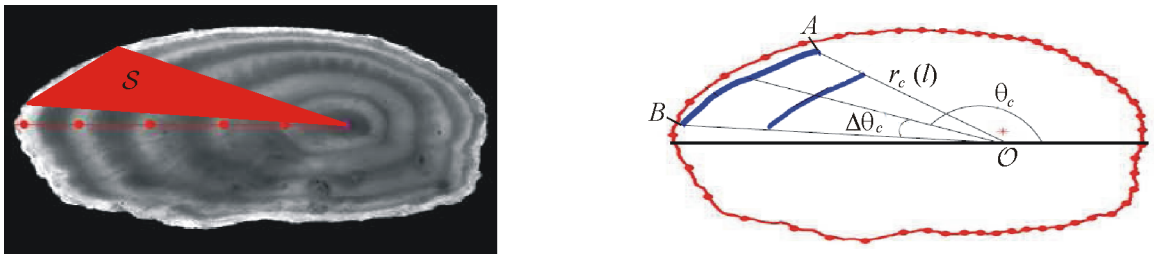


Figure 2.7: Illustration of the template-based approach aimed at extracting a 1D signal representation of the image content within a region of interest under transmitted-light (from Fablet 2006). Yearly translucent rings are indicated by markers (left). The relevant points of the region of interest are represented (right).

2.4.2 Extraction and Preprocessing of Age Structures

Most of the work developed so far regarding aging estimation systems focuses on the characterization of intensity profiles (Morison et al. 1998, Troadec et al. 2000, Fablet et al. 2003). This signal is normally obtained from the center along a predefined axis, or by calculating the median of several axes in predefined image sections, depending on the amount of contrast variation between opaque and translucent zones. The second approach, which is illustrated in Fig. 2.7), is now described.

2.4.2.1 One-dimensional Intensity Profiles

Given the otolith center $\mathcal{O}$ and an angular sector $\mathcal{S}$ the following 1D signal $i(l)$ is computed:

$$i(l) = \text{med}_\theta \left\{ I \left[\mathcal{O} + \mathcal{T}_\mathcal{S} \left(l, \theta \right) \right] \right\}. \tag{2.23}$$

where $\mathcal{T}_\mathcal{S}$ is the template model used within $\mathcal{S}$, defined by the radial coordinate points,

$$\mathcal{T}_\mathcal{S}(l, \theta) = \{ r_\mathcal{S}(l) \cos\theta_\mathcal{S}, r_\mathcal{S}(l) \sin\theta_\mathcal{S} \} \tag{2.24}$$

where $\theta_\mathcal{S}$ denotes angle value for each contour point of the external shape of the otolith, and $r_\mathcal{S}(l)$ corresponds to the radial distance calculated from the otolith center to l. The function med$(\cdot)$ is the median operator implemented on the image intensity level I for all coordinate points and along the range of θ_S, where $\theta \in (A, B)$. In other words, the one-dimensional semi-local detection of ridge and valley structures is performed within $N_\mathcal{S}$ overlapping radials, referenced by $\theta_\mathcal{S} \pm \triangle\theta_\mathcal{S}$ and the length variable $l = 0, \ldots, L$, which is characterized along the radial and reflects fish life duration (see Fig. 2.7b).

Interpretation of age information from this signal relies on the detection of ridge and valley structures corresponding to the translucent and opaque rings. In order to detect

these structures within $\mathcal{S}$, one needs to extract the meaningful extrema of $\boldsymbol{i}$. In general this is accomplished by determining the zero-crossings of derivative of $\boldsymbol{i}$, but as otolith growth decays with fish age and the gray-scale intensity values are not stationary in terms of offset, it is necessary to adapt the periodicity at which age structures appear along the radial length, and to filter out the contrast variation.

2.4.2.2 Supervised Growth Demodulation (SGD)

Demodulation operations are carried out by means of a prototype model representing the fish growth, whose parameters should be tuned appropriately to obtain periodic representations of age structures (Fablet and Josse 2005). In the supervised (or manual) mode, this process is based on the use of known otolith age data, previously obtained by an expert.

Let $\boldsymbol{i}(l) = \{i(l[0]), \ldots, i(l[N-1])\} \in \mathbb{R}^N$ be the growth-modulated intensity vector at hand, computed from the previous template-based approach. Here, the length variable $l \in [0, L_\infty)$ has a total number of N samples and a regular grid, with $l[N-1] = L_\infty$ being the full length of the radial. So, the signal is also defined as $i_n = \{i_0, \ldots, i_{N-1}\};\ n \in \mathbb{Z}$.

The main consideration underlying this scheme is that the growth is a monotonically increasing function, $l = v(t)$, whose rate decays with the time variable, t. To compensate this effect, the original samples presenting a regular grid in l-domain are transformed to those of the regular grid in t-domain (Fig. 2.8). Let $\boldsymbol{i}(v) = \{i(v(t[0])), \ldots, i(v(t[N-1]))\}$ denote such a signal. This corresponds to the sample interpolation of the original signal i_n, calculated in the linear case from

$$\boldsymbol{i}(v) = i_{n-1} + (v(t) - l_{n-1}) \frac{i_n - i_{n-1}}{l_n - l_{n-1}}, \tag{2.25}$$

where $\boldsymbol{i}(v)$ is piecewise linear for $v(t) \in [l_{n-1}, l_n)$ and $t[n] = nT$, for $n \in [0, N-1)$.

Since the true growth is unknown, the goal consists in finding an estimation $\hat{v}(t)$ that obtain the most possible periodic representation of age structures. Although several growth curves may apply to this problem, the Von Bertanlanffy growth approach (VBGF) (Beverton and Holt 1957) is, the reference for many fish applications (Chen et al. 1992, Quinn and Delriso 1999). In those cases however the observed radial length is defined for every i-th otolith l_i, and is associated to age t_i as

$$l_i = v(t_i) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma_{l,i}^2\right). \tag{2.26}$$

Then, the otolith growth model is defined by

$$v(t) = L_\infty \{1 - \exp[-k(t - t_0)]\}, \tag{2.27}$$

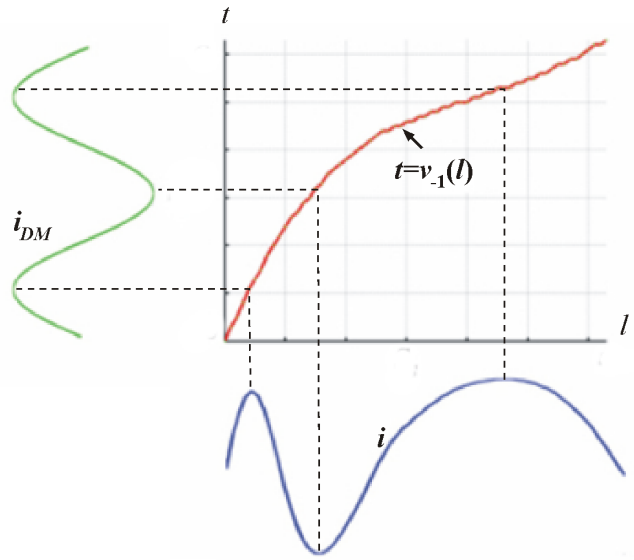


Figure 2.8: Illustration of the unwarping (or demodulation) operation. The blue trace represents the original intensity profile, i. The irregular positions in l-domain are changed to a new t-domain by means of the inverse of the otolith growth function, $t = v_{-1}(l)$, in order to obtain another profile, i_{DM}, of regular growth rings

where $\beta = (L_\infty, k, t_0)$ represents the parameters to be estimated for this purpose, where L_∞ denotes theoretical maximum length, k corresponds to the Brody growth coefficient and t_0 denotes an age for the initial fish length. The coefficient of variation of the normally distributed process error, CV_l, is assumed to be the same for all fishes. The standard deviation of the process error in length for the i-individual, $\sigma_{l,i}$ is, then, given by $\sigma_{l,i} = CV_l l_i$.

Current VBGF estimation techniques use either nonlinear *curve fitting* or *likelihood* methods to fit deterministic relationships in fish growth.

- The *Levenberg-Marquardt algorithm* (LMA)

The LMA is a very popular curve-fitting algorithm valid for estimating generic curves (Levenberg 1944, Marquardt 1963). It provides a framework to the problem of minimizing a function (in the least squares sense), generally non linear, over a space of parameters. More precisely, given a set of N empirical datum pairs of independent and dependent variables (l_N, t_N) corresponding to an unknown function, $l = v(t)$, the problem consists of optimizing the parameters β of a known growth curve model $\hat{v}(t, \beta)$ so that the sum of the squares of the deviations,

$$\varepsilon(\beta) = \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} [l_i - \hat{v}(t_i, \beta)]^2 \tag{2.28}$$

becomes minimal.

Like many other minimization algorithms, a solution is reached after an iterative procedure. To start a minimization, the user has to provide an initial guess for the parameter vector β, along with the model prototype for $\hat{v}(t, \beta)$. In each step, the parameter vector β is replaced by a new estimate $\beta + \delta$.

To determine a suitable value for δ, the functions $\hat{v}(t_i, \beta + \delta)$ are replaced by their linearizations,

$$\hat{v}(t_i, \beta + \delta) \approx \hat{v}(t_i, \beta) + J_i \delta \tag{2.29}$$

where,

$$J_i = \frac{\partial \hat{v}(t_i, \beta)}{\partial \beta} \tag{2.30}$$

is the gradient of $\hat{v}$ with respect to β, which is expressed as a row-vector.

At the minimum of $\varepsilon(\beta)$, its gradient with respect to δ will be zero. Then, the previous linear approximation gives

$$\varepsilon(\beta + \delta) \approx \sum_{i=1}^{N} (l_i - \hat{v}(t_i, \beta) - J_i \delta)^2 . \tag{2.31}$$

Or in vector notation,

$$\varepsilon(\beta + \delta) \approx \|\mathbf{l} - \hat{\mathbf{v}}(\beta) - \mathbf{J}\delta\|^2 . \tag{2.32}$$

Taking the derivative with respect to δ and equating to zero gives:

$$\left(\mathbf{J}^{\mathrm{T}}\mathbf{J}\right) \delta = \mathbf{J}^{\mathrm{T}} [\mathbf{l} - \hat{\mathbf{v}}(\beta)] \tag{2.33}$$

where $\mathbf{J}$ is the Jacobian matrix whose i-th row is J_i, and $\hat{\mathbf{v}}$ and $\mathbf{l}$ are vectors with i-th component $\hat{v}(t_i, \beta)$ and l_i, respectively. This expression is a set of linear equations which can be solved for determining δ, therefore providing the optimal increments for elements in β in a column vector.

Usually, a unitary vector $\beta^{\mathrm{T}} = (1, 1, 1)$ is normally used as an initial configuration in very simple problems. In practice, however, the algorithm converges if the initial value is already close to the final solution. In order to solve this problem, Levenberg (1944) replaced the previous equation by introducing a "damping factor",

$$\left(\mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda \mathbf{I}\right) \delta = \mathbf{J}^{\mathrm{T}} [\mathbf{l} - \hat{\mathbf{v}}(\beta)] , \tag{2.34}$$

where $\mathbf{I}$ is the identity matrix. The damping factor λ is adjusted at each iteration. If ε reduces rapidly, smaller values of λ are used. Levenberg's algorithm has the disadvantage that, for very large λ, inverting $\mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda\mathbf{I}$ causes no movement of β. For that reason, Marquardt replaced the identity matrix $\mathbf{I}$ with the diagonal of $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ in order to scale each component of the gradient so that there is larger movement along the directions where the gradient is smaller. Thus, it avoids low convergence in the direction of small gradient.

Although further proposals for the damping factor were put forward, this modification constitutes the main LMA algorithm of many software packages. The LMA method is capable of finding good solutions even if it starts far from the final minimum but for commonly used functions and reasonable starting parameters it tends to be a slightly slower than linear methods, such as the Gauss-Newton algorithm (GNA) (Fletcher 1987).

- The *likelihood* approach

The likelihood approach is similar to LMA, excepting that the least square error in equation 2.28 is replaced by the likelihood function

$$\ell = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_{l,i}} \exp\left[-\frac{\left(l_i - \hat{v}\left(t_i, \beta\right)\right)^2}{2\sigma_{l,i}^2}\right], \tag{2.35}$$

which is maximized.

Traditionally, only one age per fish is used to fit the VBGF. If multiple reads of each individual are available from different readers ($t_{i,j} : j = 1, \ldots, r$), the fish age can be approximated in three ways: a) the use of the readings of a representative reader, b) the average value or c) the median value of multiple reads. A further refinement, however, incorporates reader precision error of age into the likelihood function to improve the estimation of the growth function. In this case, a relationship between the observed and true ages can be formulated as

$$t_{i,j} = T_n + \varepsilon_{T,ij}, \quad \varepsilon_{T,ij} \sim N\left(0, \sigma_{T,i}^2\right), \tag{2.36}$$

where the standard deviation of the age-reading error for the i-th fish is $\sigma_{T,i} = CV_T T_i$ (Piner et al. 2005), being CV_T the variation coefficient of aging error. The estimated value of CV_T is calculated for multiple age reads as

$$\widehat{CV}_T = \sqrt{\frac{\sum_{n=1}^{N} CV_{t,n}^2}{N}} \tag{2.37}$$

where $CV_{t,n}$ represents age coefficient variation for the i-th fish among all readers. In this estimation, it is assumed that aging is unbiased.

Given the models for the errors observed in lengths (Eqs. 2.26 and 2.27), and estimated ages (Eq. 2.36), Cope and Punt (2007) provide a likelihood function which include the distribution of the true ages.

$$\ell = \prod_{i=1}^{N} \int \frac{1}{\sqrt{2\pi}\sigma_{l,i}} \exp\left[-\frac{\left(l_i - \hat{v}\left(t_i, \beta\right)\right)^2}{2\sigma_{l,i}^2}\right] \prod_{j=1}^{r} \frac{1}{\sqrt{2\pi}\sigma_{T,i}} \left[\exp\left(-\frac{\left(t_{ij} - T_i\right)^2}{2\sigma_{T,i}^2}\right)\right] f(T)dT \tag{2.38}$$

where $f(T)$ can be either the exponential or gamma pdf in age of the considered fish sample.

2.4.2.3 Contrast cancellation

Another operation that is generally required to improve the representation of age structures is the removal of the non uniform trending profile caused by image contrast variation. This could be implemented either on original age data or on the demodulated signals $\{\boldsymbol{i}(v)\}$ of the previous method.

- *Manual Contrast Extraction* (MCE)

In a more manual fashion, the contrast profile can be by interpolating significant points of the contrast level, c. Let $c_k = \{c_0, \ldots, c_K\}$ denote such points, where $k << T$. Then, the contrast level $\boldsymbol{i}_D$ is estimated as

$$i_D(t) = c_{k-1} + (t - t_{k-1}) \frac{c_k - c_{k-1}}{c_k - c_{k-1}}, \tag{2.39}$$

where $\boldsymbol{i}_D$ is piecewise linear within $t \in [t_{k-1}, t_k)$. As the goal is not to fit the profile but to remove the contrast level, $\boldsymbol{i}_C = \boldsymbol{i} - \boldsymbol{i}_D$ the fit error is not a relevant information in this operation.

- *Adapted growth filtering* (AGF)

Another possibility is to directly estimate the contrast component by convolving with a Gaussian kernel $\boldsymbol{g_C}$ of large width $\sigma_C^2 : \boldsymbol{i_C} = \boldsymbol{g_C} * \boldsymbol{i}$. Furthermore, this process can be also used to filter out quantization noise generated in the inner digitization of the image in the gray-scale values, so the new intensity profile signal becomes

$$\sigma_Q^2 : \boldsymbol{i_Q} = \boldsymbol{g_Q} * (\boldsymbol{i} - \boldsymbol{i_C}), \tag{2.40}$$

where $\boldsymbol{g_Q}$ is another Gaussian kernel but of much lower variance ($\sigma_Q \ll \sigma_C$).

The processed signal $\boldsymbol{i_Q}$ may offer proper conditions to determine its frequency content, so one could consider using feature extraction techniques. However, preliminary results for plaice otoliths, using the discrete cosinus transform (see subsection 3.2.4.3 in chapter 3), suggest that growth-adapted filtering alone is not likely to improve classification improvement (Fablet and Josse 2005). As such, further preprocessing of the intensity profile is required.

2.4.2.4 Peak-Based Representation (PB)

A more advanced method tries to detect the exact position of translucent and opaque zones. Fablet and Le Josse (2005) proposed a 'peak-based' representation method that calculates a version of the signal, based on the assumption that the position of rings in $\boldsymbol{i_Q}$ correspond to the regular positions $\{l_1, \ldots, l_k\}$ of local maxima, calculated from the zero crossings of the first-order derivative

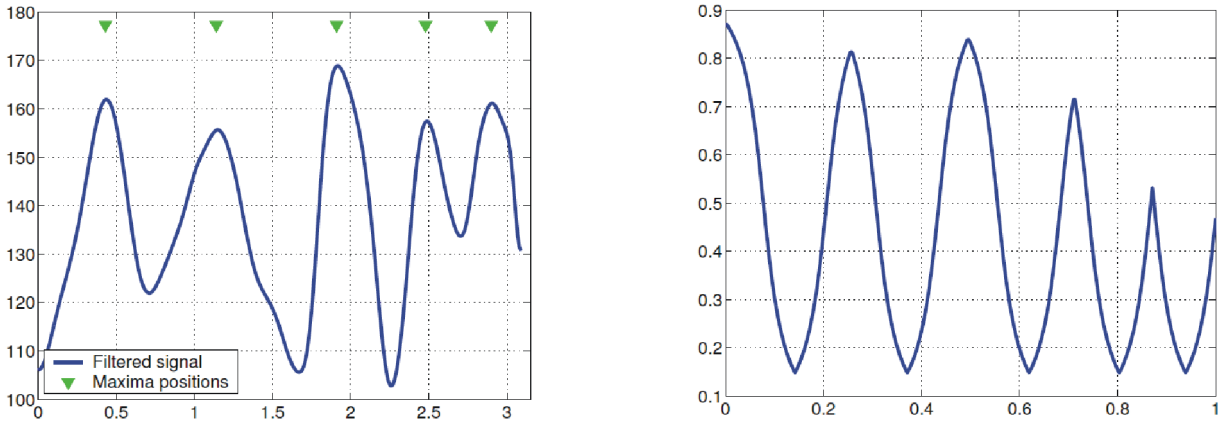
$$l_k = \left\{ l : \frac{di_Q(l)}{dl} = 0 \right\}, \tag{2.41}$$

for $k \in \mathbb{Z}$. As l_k also include the positions of local minima of i_Q, only half of the the values $l_k^* = l_{2k-1}$ should be considered. Thus, the peak-based representation $\boldsymbol{i_{PB}}$ is then defined by

$$i_{PB}(l) = \min_k \left\{ f_{k \in \{1,\ldots,K-1\}} \right\}, \tag{2.42}$$

where $\boldsymbol{f}_k = f(l - l_{k-1})$ is a kernel function centered at l_k, $\boldsymbol{f}_k \in [l_{k-1}, l_{k+1})$, being $l_0 = 0$ and $l_K = L$, respectively. Thus, the demodulated signal is changed by a synthetic function of much more accentuated oscillatory trends, constructed from k piecewise functions whose extrema positions are centered with fish year marks (Fig. 2.9). Although the original proposal used the kernel $f(u) = 2\arctan(\alpha u)/\pi$, where α is a scaling parameter, Gaussian and other trigonometric kernels are also possible. However, one important issue concerns the width that the window should have, as a bad selection may alter the true frequency structures and, subsequently, the fish age.

This configuration has proved to obtain the best classification results for plaice otoliths, using SVM's classifiers (90%). However, this is considered a very "easy-to-read" species in the fisheries industry (Fablet 2006) and the PB feature extraction methodology has not been tested in more complicated ones.



(a) (b)

Figure 2.9: illustration of the peak-based representation for the otolith image represented in Figure 2.7 (from Fablet 2006). a) Signal $i(l)$ with its extracted maxima positions (green markers), b) Associated peak-based representation $i_{PB}(l)$.

2.5 Discussion

The different methods regarding the digitization and preprocessing of otolith application data have been reviewed in this chapter. These include the manual preparation of otoliths, lightning conditions, image acquisition, noise removal and the preprocessing required to obtain computer data. While extracting contour data from binarized images is only required in fish identification applications, the preparations for the aging tasks are much more complex.

Extracting contour data points requires the combination of segmentation methods and thresholding. Since light conditions can be easily controlled in this application, the main problems arise in the form of translation and rotation effects. A possible solution is to normalize the contour profile in terms of its size, orientation, and starting point. Although the normalization problem is not the main purpose of this thesis, a new framework is proposed for this problem in chapter 5.

The contour radials form the Standard Euclidean Basis (STD) in the fish identification experiments of this work and constitute the main reference domain in the study of *Merluccius* species. Species of convex otoliths, which may require multivariate representation, are not considered, as only feature extraction techniques valid for one-dimensional signals have been developed.

Current fish aging methodology, on the other hand, is based on transmitted-light mode to increase contrast variation between the translucent and opaque ring structures observed in the gray-scale otolith image. While the detection of the nucleus was extensively covered

in the work of Cao and Fablet (2006), fish growth estimation has become the main subject of research in recent decades.

The main approach used in the generation of biological growth models is based on the Von Bertanlanffy function. Different solutions to this problem propose the use of different error cost functions, including among other parameters: seasonal variations (Pitcher and McDonald 1973), individual variation (Wang and Thomas 1995, Pilling et al. 2002), stochastic environmental fluctuations (Prajneshu and Venugopalan 1999) or reader age precision error (Cope and Punt 2007). In our opinion, however, the extent to which these approaches can be applied is questionable. The determination of a generalized single set of growth parameters for all the specimens suggests that all specimens obey the same growth law, something very hard to assume in practice, if we consider the deterministic role that genetic factors play in individual specimens.

The decreasing trending profile of gray-scale values (Fig. 1.6) suggests the existence of an image contrast whose magnitude varies along the radial. Thus, if σ_C is not properly tuned at different positions of the intensity profile, the signal may appear considerably distorted, specially at the tail of the profile where specimens of large age concentrate the structures of interest and the profile slope is steep. For the same reason, σ_Q must be large enough to avoid the effect of quantization noise when computing derivatives, as well as low enough to maintain the oscillatory components of interest. On the other hand, it is also very hard to assume that the intensity profile can be literally changed for a synthetic signal. In this sense, as year marks are represented by detected zero derivatives the role of feature extraction in this framework is questionable.

In fact, the method has demonstrated to be robust only for plaice otoliths but has not been extended to more complicated ring structures present in other species. This thesis introduces a new approach of automatic aging system where growth functions are optimally estimated individually for each single otolith. This approach tries to be consistent with the assumption that genetic factors, combined with environmental conditions of fish habitat, may influence every single fish differently. This is the reason why the estimation of VBGF parameters needs to be tuned individually for each fish, having a single an independent β_i for every intensity profile i.

Part II

On Feature Extraction

Chapter 3

Feature Extraction

3.1 Introduction

Although te field of pattern recognition has been more concerned with signal detection, classification, regression and machine learning, among other applications (Devijer and Kittler 1982, Fukunaga 1990, McLachlan 1992); the topic of *feature extraction* has experienced significant development recently with the implementation of a wide range of automatic algorithms, all of which serve to determine important information from data.

This chapter deals with these algorithms. Traditionally, although practical applications have been associated to digital data compression/decompression systems (commonly known as codecs in sound, image and video applications), pattern recognition, however, is more concerned with the problem of making classification information more "visible". In this sense, both problems are regarded as the *dimensionality reduction* problem, which is defined as follows,

Definition. 3.1 *Given a feature space* $x \in \mathbb{R}^N$*, find a mapping* $y = f(x) : \mathbb{R}^N \to \mathbb{R}^K$*, where* $K<N$*, such that the transformed feature vector* $y \in \mathbb{R}^K$ *preserves* (*most of*) *the* useful *information or structure in* $\mathbb{R}^N$.

If the objective is to make data more compact, large energy components are representative quality factors that can be found in the data whereas for the case of obtaining class information, discriminatory components are the perfect choice (Saito and Coifman 1995).

The latter group is what we are interested in. Finding different and reduced data sets of features that more accurately explain the nature of the problem is crucial in the presented applications: otolith-based fish identification and age estimation. In general, the Standard Euclidean Basis (STD) is insufficient to attain this goal. Some potential reasons are: a) noisy

or irrelevant features will habitually impact negatively on accuracy, as things tend to look more similar when the dimensionality increases (on average); and b) the retrieval time of classification results increases dramatically with the size of the feature vector used.

In this thesis, we have exploited the *Best-basis paradigm* (Coifman and Wickerhauser 1992, Meyer 1993), which follows these three main steps:

1. select a "best" basis (or coordinate system of features) for the problem at hand from a *library* (a fixed yet flexible set of bases consisting of wavelet packets, local trigonometric bases or any type of wavelets).

2. sort the coordinates by "importance" and discard the useless coordinates.

3. use the surviving coordinates to develop classification tasks.

Although this procedure may be seen as a filtering process, actually it is a mapping $\boldsymbol{y} = \mathcal{F}(\boldsymbol{x})$. That is, the filtering process is guided by an objective function, which is maximized or minimized depending on the purpose at hand, in order to select the useful components. Clearly, what is useful is defined in the objective function. This makes the Best-Basis paradigm appropriate for clustering problems.

3.1.1 Why Wavelets

In general, data may be transformed to find better forms for describing its characteristics. Fourier descriptors have been the common tools for this operation in many otolith-based fish identification and aging tasks. When combined with statistical analysis such as PCA or LDA they constitute a powerful tool for characterizing morphological differences of many fish species. Similarly, they allow inter-specific analysis for species recognition purposes.

When considering the frequency content, Fourier-based methods are only appropriate if the oscillatory components remain within the whole time interval because the integral is defined over the entire range (Mallat and Hwang 1992). However, for either the otolith contour or the one-dimensional intensity profile, both of which come in all forms of edges, spikes, and transients (Lombarte and Castellón 1991), basis functions able to fit to such irregularities and detect their position are preferable. Wavelets and their relatives provide such a flexible coordinate system, since they allow these kind of features to be captured in a computationally efficient manner (Mallat 1989, Parisi-Baradad et al. 2005).

This structural approach forms a description language that can be understood as follows: the words (the elementary building blocks) are the functions that characterize the basis vectors of the wavelets and their relatives. A collection of words defines a dictionary which

corresponds to the set of bases specified by their time-frequency location characteristics (Fig. 3.1). A library is a collection of dictionaries (or books), each containing a necessary language to describe things, in the same way that each mother wavelet shape provides a basis function of reference to explain irregularities. And, finally, the necessary grammar for describing certain types of things (or class of irregularities) is defined by the words (or selected basis vectors) of one dictionary from the library.

This paradigm provides an array of tools that bridges between the "two extremes", the Standard Euclidean Basis (or position of irregularities) and the Fourier Basis (frequency information); and leads to a vastly more efficient form of representation and processing of signals, compared with strategies confined to a single basis. If one succeeds in extracting the correct words from the books of the library, the problem of pattern recognition is simplified and solved efficiently.

3.1.2 Overview of this Chapter

Several methods for data analysis will be reviewed in this chapter. They include 1) *descriptive* tools: Fourier, trigonometric and wavelet transforms; and 2) *analytical* tools: component analysis and discriminant analysis.

The focus of our discussion is the wavelet transform, and specially, the *discrete wavelet packet transform* (DWPT) withs its library variants (section 3.2). Section 3.3 introduces the Best-Basis paradigm. The focus is on the selection of features for classification purposes. This will lead to the Local Discriminant Basis (LDB) algorithm, the main tool under consideration in this thesis. We also provide graphical tools to determine some useful properties in data. In particular, we explain how the location and frequency properties can be determined from the basis functions (section 3.4). Drawbacks and advantages related to the different feature extraction schemes will be discussed in section 3.5.

3.2 On Description Methods for Feature Extraction

Most often, information is presented in the form of discrete sequences $x[n]$ obtained from a finite resolution device (or analog-to-digital converter) which averages and samples an analog signal. That is,

$$x(t) \approx \hat{x}(t) = \sum_{n=0}^{N-1} x[n] \mathcal{X}(t-n) \tag{3.1}$$

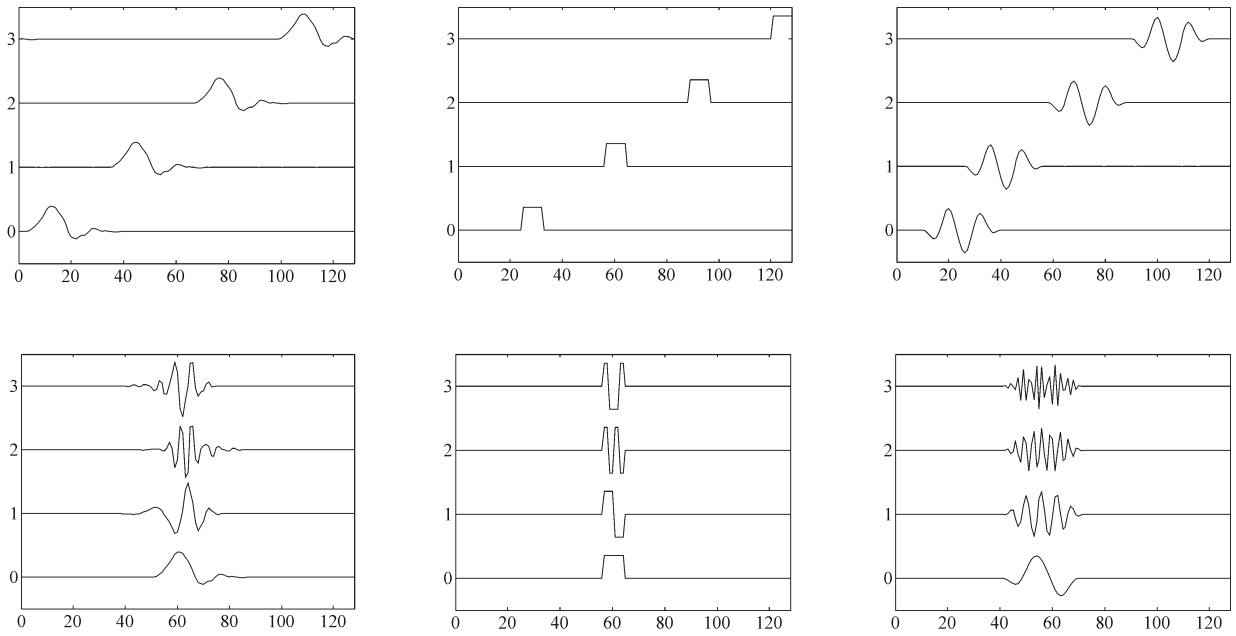


Figure 3.1: Examples of basis functions available for feature description. From left to right: the Daubechies 4 wavelet packet basis, the Haar basis and the discrete sine basis. Top row: Translated basis functions. Bottom row: Basis functions with different frequencies. Horizontal axes indicate time in this figure.

where $\mathcal{X}(t)$ is an averaging function representing the analog-to-digital converter. For example, a CCD camera filters the light intensity by the optics and each photo-receptor averages the input light over its support. If the sampling distance is N^{-1}, we need to associate to $\mathcal{X}(t)$ a function approximated at scale N^{-1}, so that $\hat{x}[n] \equiv \hat{x}(t) = \hat{x}(nN^{-1})$ for $n = 0, \ldots, N-1$.

In general, the selection of $\mathcal{X}(t)$ requires some knowledge about $x(t)$. If the implicit choice $\hat{x}[n] = x[n]$ is assumed, coefficient errors are restricted to the Shannon's theorem. Possible choices are $\mathcal{X}(t) = 1$ for $nN^{-1} \leq t \leq (n+1)N^{-1}$ and $\mathcal{X}(t) = 0$ if not, all of which amounts to a piecewise constant approximation of $x(t)$, or $\mathcal{X}(t) = (\sin t)/t$ for band-limited signals. It is important to note that this discretization is made prior to the description algorithm and that the choice of $\mathcal{X}(t)$ is completely independent of the basis functions and any other parameters.

Whether approximated or not, the notation $x[n]$ will be used to denote discrete input signals, from now on. Habitually, a dyadic length $N(= 2^{n_0})$ is used for developing fast algorithms (Beylkin 1993). We focus our attention on signals extended periodically beyond the interval $[0, N-1)$, $x[n] \equiv x[N+n]$, which are appropriate for the analysis of otolith contour data. As for the intensity profile of aging applications, however, border effects can be reduced by considering symmetric extensions of the opposite sign. That is, $x =$

$\{x[0], \ldots, x[N-1], -x[N-1], \ldots, -x[0]\}$.

3.2.1 Fourier Descriptors

The *Discrete Fourier Transform* (DFT) decomposes an aperiodic signal into a sum of complex exponential terms $e^{i\omega}$. Basically, it changes the time (or space) representations of signals into a representation of the angular frequency ω, commonly known as the frequency domain.

The DFT is widely employed in signal processing and related fields to analyze the frequencies contained in digital signals. Proof of the importance of this operation is the fact that the term FFT (*Fast Fourier transform*) is often used to mean DFT in colloquial settings, even though this term refers exactly to a family of algorithms that are computationally efficient (Good 1958, Cooley and Tukey 1965, Bruun 1978).

More formally, let $\boldsymbol{x} \equiv \{x[0], \ldots, x[N-1]\}$ be a sequence of N samples. This sequence is transformed into another sequence of N complex numbers $\{X[0], \ldots, X[N-1]\}$ according to the following expression of the DFT,

$$\text{DFT}: \quad X[k] = \sum_{n=0}^{N-1} x[n] \exp\left[-i\frac{2\pi}{N}kn\right] \quad k = 0, \ldots, N-1. \tag{3.2}$$

where the exponential term constitutes the basis vectors. The *inverse discrete Fourier transform* (IDFT) is given by

$$\text{IDFT}: \quad x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k] \exp\left[i\frac{2\pi}{N}kn\right], \tag{3.3}$$

and can be used to reconstruct the whole signal, or part of its components. A simple generalization of DFT is that the complex numbers $X[k]$ represent the amplitude and phase features corresponding to the sinusoidal components that makes $x[n]$, and whose frequency $\omega = \frac{2\pi}{N}k$ is indexed by k. The equations make use of the Euler notation to express this principle because sinusoids are more efficiently manipulated by means of complex exponentials. However, by writing $X[k]$ in polar form the sinusoid amplitude A_k/N and phase φ_k can be derived, respectively, as

$$A[k] = |X_k| = \sqrt{\text{Re}^2 X[k] + \text{Im}^2 X[k]}, \tag{3.4}$$

and

$$\varphi[k] = \arctan\left(\frac{\text{Im } X[k]}{\text{Re } X[k]}\right). \tag{3.5}$$

The normalization factors multiplying the DFT and IDFT, here 1 and $1/N$, and the signs in the exponent are merely conventions, and are used differently depending on the application type. In this sense, the equations for the the Fourier series are similar except that the term $1/N$ is included in the expression $X[k]$ instead of the expression $x[n]$. A normalization of $1/\sqrt{N}$ makes the transform unitary, offering some theoretical advantages, but the previous one is often more practical in terms of signal interpretation to perform the scaling all at once, just as the term $A[k]/N$ above. Additionally, the convention of a negative sign in the exponent is preferable because in this case it means that $X[k]$ is the amplitude of a positive frequency $\omega = 2\pi k/N$. In practice, however, the coefficient values belonging to the negative frequencies equal those of the positive ones, and the analysis thus will be restricted to $N/2$ coefficients when N is even, or (N+1)/2 coefficients if N is odd.

The coefficients can be used for different purposes, including: spectral analysis, data compression, interpolation, or the multiplication of polynomial and or large integers; among others. When the inverse operation (IDFT) is restricted to some finite number of coefficients, filtering operations are implemented over the initial signal.

3.2.2 STFT. The Short-Time Fourier Transform

The discrete *short-time Fourier transform* (STFT) is a Fourier-related transform that can determine the variation of frequency and phase components of a signal as it changes over time. This is possible because the coefficients are evaluated at a given position (or time), u, instead of the whole signal time support. Therefore, the STFT transform is given by

$$\text{STFT}: \quad X[u,k] = \sum_{n=0}^{N-1} x[n]\, g[n-u] \exp\left[-i\frac{2\pi}{N}kn\right]. \tag{3.6}$$

where $g[n-u]$ corresponds to a symmetric window of period N centered at position u and modulated in frequency by k, where $0 \leq u < N$ and $0 \leq k < N$. Similar to the IDFT, the original signal is reconstructed by the corresponding inverse short-time Fourier Transform (ISTFT) as

$$\text{ISTFT}: \quad x[n] = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{k=0}^{N-1} X[u,k]\, g[n-u] \exp\left[i\frac{2\pi}{N}kn\right] \tag{3.7}$$

In principle, this type of transform has the potential advantage that it can capture frequency variations in time. The slice of information provided by the atom $\phi = g[n-u]\, e^{i\frac{2\pi}{N}kn}$ in Eq. 3.6, corresponds to a region in a time-frequency plane (n, ω) whose location and width depends on the time-frequency spread of ϕ. Since

$$\|\phi\|^2 = \sum_{n=0}^{N-1} |\phi[n]|^2 = 1, \tag{3.8}$$

$|\phi(n)|^2$ can be thought of as a probability distribution centered at

$$u = \sum_{n=0}^{N-1} n\,|\phi[n]|^2\,. \tag{3.9}$$

The spread around u is measured by the variance

$$\sigma_n^2 = \sum_{n=0}^{N-1} [n-u]^2\,|\phi[n]|^2\,. \tag{3.10}$$

The Plancherel formula proves that $\sum_{k=0}^{N-1} \left|\dot{\phi}[k]\right|^2 = 2\pi\,\|\phi\|^2$, where $\dot{\phi}$ denotes DFT of ϕ (Plancherel and Leffler 1910). Considering only the analytical part, the center frequency of $\hat{\phi}$ is therefore defined by

$$\xi = \sum_{k=0}^{N/2} k \left|\dot{\phi}[k]\right|^2, \tag{3.11}$$

where if N is even (or $(N+1)/2$ if N is odd). The spread around ξ is

$$\sigma_\omega^2 = \sum_{k=0}^{N/2} [k-\xi]^2 \left|\dot{\phi}[k]\right|^2. \tag{3.12}$$

Habitually, the resolution of time-frequency components is expressed by the area of a time-frequency box (sometimes referred to as Heisenberg box) centered at (u, ξ), and whose area is established by σ_u and σ_ω, respectively (Fig. 3.2). Since g is even, the size of the box is independent of u and ξ, and it remains fixed. This means that all atoms in the STFT have the same time-bandwidth resolution across the time-frequency plane. The uncertainty principle excludes the possibility of having arbitrarily high resolutions, since in the best case (a Gaussian window) the time-bandwidth product $\sigma_u\sigma_\omega$ can not be lower than 1/2 (Wheeler and Zureck 1983).

When using a discrete grid, STFTs can be highly redundant: small increments of u and ω cause atoms to overlap so that $x(n)$ can not be recovered exactly by means of ISTFT. The solution to this problem resorts to using a frame such that the basis functions constitute an orthonormal basis. Unfortunately, this is possible only if $g(\cdot)$ is badly localized in either time or frequency, which normally is the reason why the STFT is oversampled. Thus, the use of

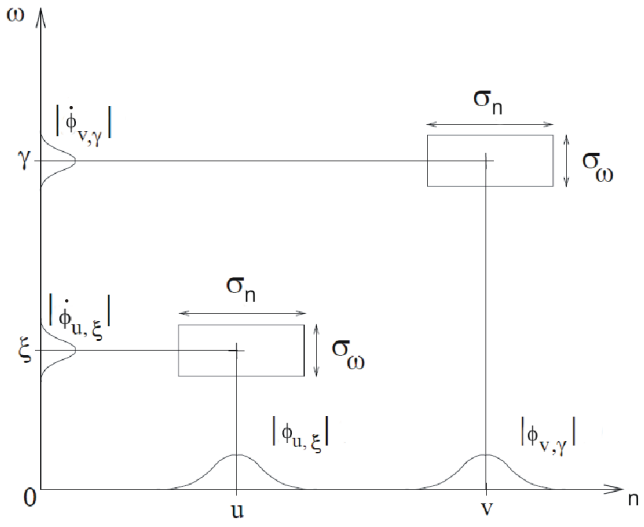


Figure 3.2: Heisenberg boxes of two windowed Fourier atoms $\phi_{u,\xi}$ and $\phi_{v,\gamma}$ (from Mallat 1990).

STFT is limited to audio monitoring tools, such as spectrograms, which are implemented by calculating the value $|X[u,k]|^2$.

3.2.3 Wavelet Transforms

By varying the window width, time resolution can be traded for resolution in frequency or vice-versa. In this sense, some short basis functions could be used to isolate discontinuities of high frequency, while the remaining ones could be long functions to develop spectral analysis towards the lower frequencies. An intuitive way to do this is by having both short high frequency and long low frequency functions available. This is exactly what is achieved with the *wavelet transform* (Daubechies 1988, Mallat 1989, Daubechies 1992, Mallat and Hwang 1992, Meyer 1990).

In a general sense, the **Continuous Wavelet Transform** (CWT) calculate the coefficients as inner products of $\boldsymbol{x}$ against a family of dilated and translated wavelets. Its general expression is given by

$$\text{CWT}: \quad C_{s,u} = \int x(t)\, \psi^*_{u,s}(t)\, dt, \tag{3.13}$$

where the dilated and translated versions are obtained from a prototype wavelet (or mother function) $\psi(t)$, corresponding to a zero average high-pass filter ($\sum_n \psi[n] = 0$), given by the model

$$\psi_{u,s}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-u}{s}\right) \tag{3.14}$$

(the asterisk stands for complex conjugate). The time position, spread and the frequency characteristics of the mother wavelet $\psi(t)$ (including center frequency and bandwidth) are obtained in a similar fashion to the short-time Fourier transform (just change ϕ by ψ from equations 3.9 to 3.12). The wavelet transform can also be written as a convolution product

$$\text{CWT}: \quad C(s,u) = \boldsymbol{x} * \overline{\boldsymbol{\psi}}_s(u), \tag{3.15}$$

where $\overline{\psi}_s(t) = s^{-1/2}\psi^*(-t/s)$.

All the parameters, including the time variable t, are continuous. The different possibilities of discretization constitute the variants of this methodology and the *Wavelet Series* theory. As an example, the **Discrete-Time Wavelet Transform** (DTWT) is defined as the CWT of a discrete sequence $x[n] \equiv x(nT)$, for T=1

$$\text{DTWT}: \quad C(s,u) = \sum_n x[n]\,\psi^*_{u,s}[n]. \tag{3.16}$$

All the versions constructed from the mother wavelet are also high-pass and share the same properties but have different location and spread in the time-frequency plane. Since the Fourier transform of $\overline{\psi}_s(t)$ is $\dot{\overline{\psi}}_s(\omega) = \sqrt{s}\dot{\psi}^*(s\omega)$, it holds that increasing the dilation parameter, s, reduces the center frequency of the basis function and improves its bandwidth resolution but, in contrast, it worsens time resolution (see Fig. 3.3). The center frequency, ξ, and the bandwidth, σ_ω, for each basis function become $\xi(s) = \xi/s$ and $\sigma_\omega(s) = \sigma_\omega/s$ respectively, whereas the time resolution σ_n is calculated as $\sigma_n(s) = s\sigma_n$.

The exact recovering of $x[n]$ using the CWT coefficients is less evident. For this goal, the different wavelet atoms must meet certain properties:

- **Compact support**: All together, the disjoint basis functions must cover the whole time-frequency support in the Heisenberg plane without redundancy.
- **Orthogonality**. The basis vectors must form an *orthonormal* basis of the space $x(t)$ belongs to.
- **Scaling Function**. An additional basis function is necessary to represent the lowest frequency components of $x(t)$ which are lost with the use of ψ.

Compact support means that the extension of all basis functions must cover the whole time-frequency plane (which may have different time u_0 and frequency ξ/s location) without

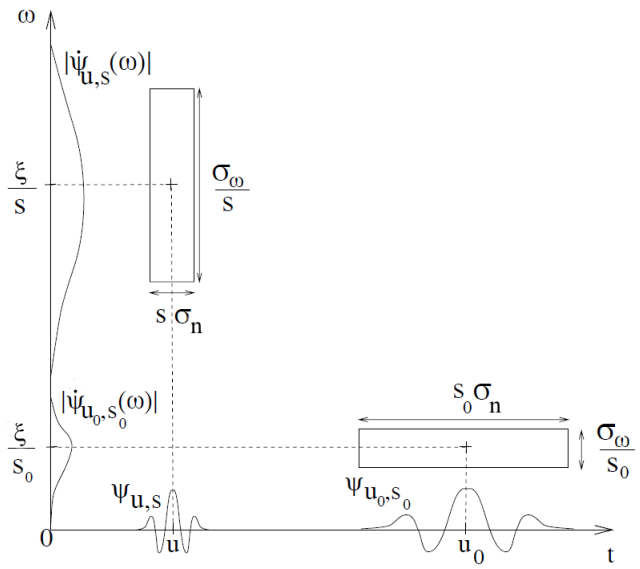


Figure 3.3: Heisenberg boxes of two wavelets. Smaller scales decrease the time spread but increase the frequency support, which is shifted towards the higher frequencies (from Mallat 1990).

overlapping. Additionally, perpendicularity must exist through the basis functions of different position and scales. Thus, not all the coefficients $C(s,u)$ are necessary for the exact reconstruction of $x[n]$, a priori.

Daubechies (1992) proved that compactly supported orthogonal wavelets can be obtained if the time-scale plane is sampled on a "dyadic" grid ($s = 2^j$ and $u = k2^j$ for $j,k \in \mathbb{Z}$). This has led to the **Wavelet Series** (WS) framework which describes the different methodologies by which the CWT is discretized and the perfect reconstruction of signals is ensured, just as in any other standard description method. A usual definition is

$$C[j,k] = \text{CWT}\left\{x(t)\,; s = 2^j, u = k2^j\right\} \quad \text{for} \quad j,k = \in \mathbb{Z}. \tag{3.17}$$

The wavelets are in this case $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$ and the recovering of $x[n]$ is formalized as

$$x(t) = \sum_j \sum_k C[j,k]\,\tilde{\psi}_{j,k}(t)\,. \tag{3.18}$$

where $\tilde{\psi}(t)$denotes *synthesis* wavelets.

Finally, the original signal $x(t)$ can not be exactly recovered using orthogonal wavelets of only high-pass characteristics, since that would destroy the offset and the low-frequency components. Mallat (1989) showed that the exact recovering of the original signal $x(t)$ is

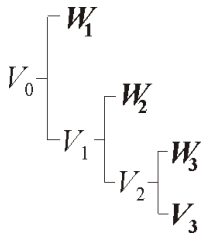


Figure 3.4: A decomposition of V_0 into mutually orthogonal spaces using the wavelet transform (with $J = 3$). The symbols in bold font represent the subspaces kept intact by the discrete wavelet transform.

possible if a complementary function $\phi(t)$ is used to reconstruct the complementary information high-pass orthonormal wavelets $\psi_{j,k}(t)$ need. This led to the *multi-resolution theory* which nowadays constitutes one of the main frameworks for developing fast algorithms for the discrete wavelet transform and filter design (see Appendix A).

Let V_0 represent the signal space $x(t)$ belongs to. Assume a function exists in this space $\phi(t) \in V_0$, such that the set $\{\phi_{j,k}(t)\}_{j,k\in\mathbb{Z}}$ forms a basis V_j representing the domain of all band-limited signals of frequency interval $(-2^{-j}\pi, 2^{-j}\pi)$. The function $\phi(t)$ is defined in a "dyadic wavelet style" and, very often, is referred to as the *scaling function* since the frequency interval of $x(t)$ is reduced half-band in every octave j. Similarly, call W_j the spaces corresponding to the wavelet set $\{\psi_{j,k}(t)\}$.

Then, the basis functions $\psi_{j,k} \in W_j$ will form an orthonormal basis of $\phi_{j,k} \in V_j$ in V_{j-1}, if the perpendicularity condition $\phi(t-k) \perp \psi(t-k)$ holds and the frequency interval of W_j complements that of V_j, which are $(-2^{-j+1}\pi, -2^{-j}\pi) \cup (2^{-j}\pi, 2^{-j+1}\pi)$. To ensure these properties the relations

$$\begin{aligned} \langle \phi(x-l), \phi(x-k) \rangle &= \delta_{kl} \\ \langle \psi(x-l), \psi(x-k) \rangle &= \delta_{kl} \\ \langle \phi(x-l), \psi(x-k) \rangle &= 0 \end{aligned} , \tag{3.19}$$

must hold (here, δ_{kl} denotes the Kronecker delta at time position $k = l$). W_j will then represent the bandpass functions in V_j that makes $x(t)$ in V_{j-1},

$$V_{j-1} = V_j \cup W_j, \tag{3.20}$$

where $V_j \subset V_{j-1}$ and $W_j \subset V_{j-1}$ (Fig. 3.4).

Eq. 3.19 establish the conditions for perfect reconstruction and lead to the two-scale difference equations

$$\frac{1}{\sqrt{2}}\phi\left(t/2\right)=\sum_{n}h_{0}\left[n\right]\phi\left(t-n\right)=\boldsymbol{h}_{0}*\boldsymbol{\phi}, \tag{3.21}$$

$$\frac{1}{\sqrt{2}}\psi\left(t/2\right)=\sum_{n}h_{1}\left[n\right]\phi\left(t-n\right)=\boldsymbol{h}_{1}*\boldsymbol{\phi}, \tag{3.22}$$

where $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ are the finite-length impulsional responses (FIR) corresponding to the low-pass $\boldsymbol{\phi}$ and high-pass $\boldsymbol{\psi}$ basis functions, respectively.

The design of these filters is a major topic of research in wavelet analysis (Vetterli 1992), and with their corresponding synthesis filters, $\overline{h}_0$ and $\overline{h}_1$ (their reversed versions in the orthogonal case), form the two-channel filter-bank necessary for the fast computation of the **Discrete Wavelet Transform** (DWT). For this reason, they are sometimes referred to as *quadrature mirror filters* (QMF)

Definition. 3.2 *A Fast DWT algorithm is computed as*

$$\mathrm{DWT}\left\{x\left[n\right];2^{j},k2^{j}\right\}=d_{j}\left[k\right]=\sum_{n}x\left[n\right]h_{1}^{(j)}\left[n-2^{j}k\right] \tag{3.23}$$

$$a_{J}\left[k\right]=\sum_{n}x\left[n\right]h_{0}^{(J)}\left[n-2^{J}k\right], \tag{3.24}$$

where $h_0\left[n\right]=\left\langle 2^{-1/2}\phi\left(t/2\right),\phi\left(t-n\right)\right\rangle$ *and* $h_1\left[n\right]=\left\langle 2^{-1/2}\psi\left(t/2\right),\phi\left(t-n\right)\right\rangle$ *denote low-pass and high-pass half-band FIR filters, and at each octave* j *the filter sequences are recomputed as*

$$\begin{array}{c} h_{1}^{(1)}\left[n\right]=h_{1}\left[n\right],\quad h_{0}^{(1)}\left[n\right]=h_{0}\left[n\right], \\ h_{1}^{(j+1)}\left[n\right]=\sum_{l}h_{1}^{(j)}\left[l\right]h_{1}\left[n-2l\right], \\ h_{0}^{(j+1)}\left[n\right]=\sum_{l}h_{0}^{(j)}\left[l\right]h_{0}\left[n-2l\right]. \end{array} \tag{3.25}$$

The terms $a_j\left[k\right]$ and $d_j\left[k\right]$ denote "approximation" and "detail" coefficients, respectively. Although the filter sequences should be calculated at each scale, the decomposition process can be organized in a way that the original structure of $\boldsymbol{h}_0$ and $\boldsymbol{h}_1$ is maintained throughout the dyadic scales.

Fig 3.5 illustrates this process. First, $a_0\left[n\right]$ is initialized to some consistent discretization of the input $\hat{x}\left[n\right]=x\left(nN^{-1}\right)$. In this sense, it is possible to show that if the implicit choice $\hat{x}\left[n\right]=x\left[n\right]$ is taken, errors can come from the distance of the scale function ϕ to an ideal low-pass filter, at various scales and "details" and, for this reason, Abry and Flandrin (1994) proposed using the approximation

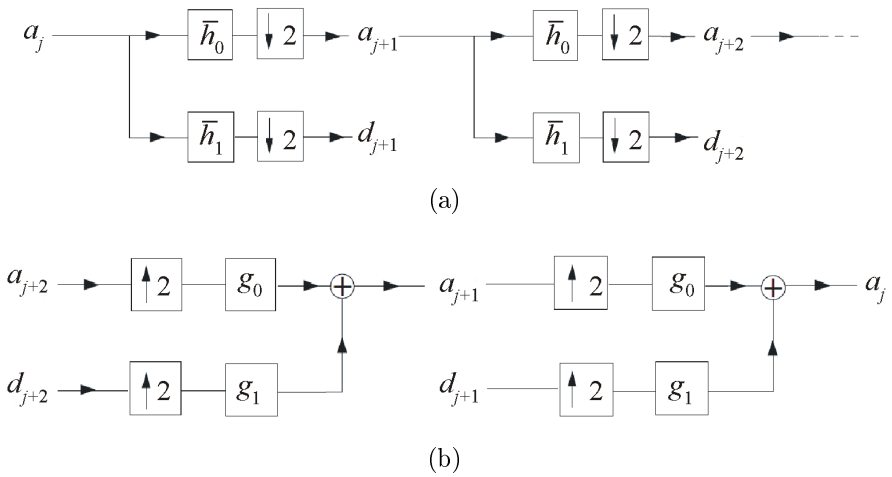


Figure 3.5: Fast discrete wavelet transform computation (from Mallat 1990); (a) A fast wavelet transform is computed with a cascade of filterings with $\overline{\boldsymbol{h}}_{0'}$ and $\overline{\boldsymbol{h}}_1$ followed by a factor 2 sub-sampling; (b) A fast inverse wavelet transform reconstructs progressively each $\boldsymbol{a}_j$ by inserting zeros between samples of $\boldsymbol{a}_{j+1}$ and $\boldsymbol{d}_{j+1}$, filtering with $\boldsymbol{g}_0$ and $\boldsymbol{g}_1$, and adding both sequences.

$$\hat{x}\,[n] = \sum_k x\,[k]\,\phi\,(k-n)\,. \tag{3.26}$$

This initialization is also seen as a prefiltering that cancels possible aliasing generated during the discretization of $x\,(t)$.

Let denote $\overline{x}\,[n] = x\,[-n]$ and

$$\check{x}\,[n] = \begin{cases} x\,[k]\,, & \text{if } n = 2k \\ 0, & \text{if } n = 2k+1 \end{cases}. \tag{3.27}$$

Then, at the decomposition we have

$$\begin{array}{l} a_{j+1}\,[k] = \sum_n a_j\,[n]\,\overline{h}_0\,[n-2k] = \boldsymbol{a}_j * \overline{\boldsymbol{h}}_0\,[2k]\,, \\ d_{j+1}\,[k] = \sum_n a_j\,[n]\,\overline{h}_1\,[n-2k] = \boldsymbol{a}_j * \overline{\boldsymbol{h}}_1\,[2k]\,. \end{array} \tag{3.28}$$

At the reconstruction we use up-samplings and convolutions

$$\begin{aligned} a_j\,[k] &= \sum_n g_0\,[k-2n]\,a_{j+1}\,[n] + \sum_n g_1\,[k-2n]\,d_{j+1}\,[n] \\ &= \check{\boldsymbol{a}}_{j+1} * \boldsymbol{g}_0\,[k] + \check{\boldsymbol{d}}_{j+1} * \boldsymbol{g}_1\,[k] \end{aligned} \tag{3.29}$$

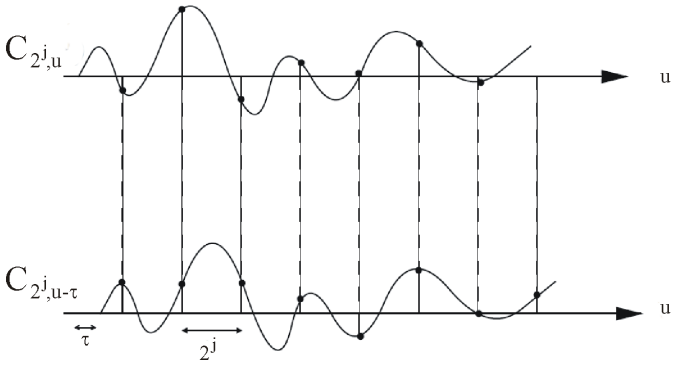


Figure 3.6: Variance translation in the computation of DWT (from Mallat 1990). If $x_\tau(t) = x(t-\tau)$, uniformly sampling $C_{2^j,u}$ and $C_{2^j,u-\tau}$ at $u = n2^j$ may yield very different values if $\tau \neq n2^j$.

The DWT is the only wavelet transform that can reproduce $x[n]$ perfectly by its inverse, except for round-off errors. The discrete inverse wavelet transform (IDWT) is given by

$$x[n] = \sum_{j=1}^{J} \sum_{k \in \mathbb{Z}} d_j[k]\, g_1^{(j)}\left[n - 2^j k\right] + \sum_{k \in \mathbb{Z}} a_J[k]\, g_0^{(J)}\left[n - 2^J k\right]. \tag{3.30}$$

where $g_0[n]$ and $g_1[n]$ denote sequences for *synthesis* filters.

This algorithm computes a total of $N/2+N/4+\cdots+N/2^{(J-1)}+N/2^J = N$ coefficients. In keeping with the critical sampling, j is restricted to $1 \geq j \geq J$ (with $J = \log_2 N$) so that the sampling rate of the coefficients is always less than the original signal. While $g_0[n] = h_0[n]$ and $g_1[n] = h_1[n]$ in the orthogonal case, these are not necessarily true in the biorthogonal case.

As opposed to complex exponentials, wavelet bases do not provide easy access to phase information. Furthermore, the sub-sampling operation of the filter-bank algorithm destroys translation invariance. Let $x_\tau(t) = x(t-\tau)$ denote the translated input signal. Indeed, although $\mathrm{CWT}\{x_\tau(t); s, u\} = C(s, u-\tau)$ is time invariant, in the DWT this property would be maintained only for $\tau = n2^j$ (see Fig. 3.6) . This situation however is rarely given in practice since the translations caused by the acquisition equipment are very difficult to predict. Additionally, the sampling rate varies at different scales for a normally fixed value of τ.

The *à trous* algorithm, introduced by Holschneider et al. (1989), circumvents this problem at the cost of increasing computational resources. This type of wavelet transform maintains the length of the coefficient vectors in all octaves and increases, instead, the length of the filters by inserting zeros between samples (see section B).

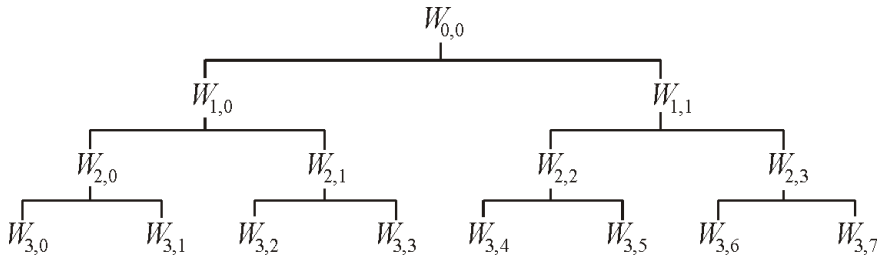


Figure 3.7: A decomposition of $W_{0,0}$ into the tree-structured subspaces using the wavelet packet transform (with $J = 3$)

3.2.3.1 Discrete Wavelet Packet Transform (DWPT)

DWT only partitions the frequency axis towards the low frequencies, whereas the components of the high frequencies (the details) are kept intact (see Fig. 3.4). Although such components may suffice in some analytical problems, in much more complicated signal patterns it may be interesting to look for differences towards the high-frequency bands.

Instead of partitioning only the approximation spaces V_j to construct detail spaces W_j, the *Discrete Wavelet Packet Transform* (DWPT) (Coifman and Meyer 1989a,b, Meyer 1993, Wickerhauser 1995) extends the partitioning to the detail spaces. They are sometimes referred to as *lifting* wavelets given their equivalence to the lifting scheme presented by Sweldens (1996, 1997).

This recursive splitting of vector spaces is represented in a binary tree (Fig. 3.7). Let $W_{j,k}$ denote any node (j,k) of this tree. The two-scale difference equation, (3.21) and (3.22), become

$$\psi_{j+1}^{2k}(t) = \sum_l h_0[l]\,\psi_j^k\left(t - 2^j l\right) \tag{3.31}$$

and

$$\psi_{j+1}^{2k+1}(t) = \sum_l h_1[l]\,\psi_j^k\left(t - 2^j l\right) \tag{3.32}$$

where, at the root, ψ_j^0 is initialized as

$$\psi_j^0(t) = \frac{1}{\sqrt{2^j}}\phi\left(2^{-j}t\right). \tag{3.33}$$

Then, for any node (j,k) the DWPT coefficients are calculated as

$$c_j^k[l] = \left\langle x(t), \psi_j^k\left(t - 2^j l\right)\right\rangle. \tag{3.34}$$

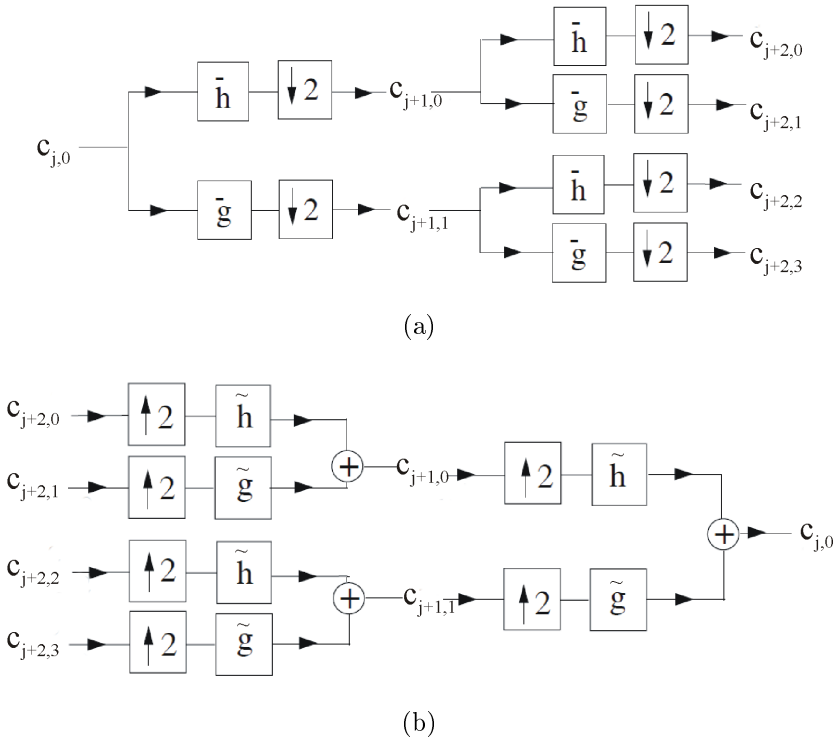


Figure 3.8: Fast DWPT algorithm (from Mallat 1990); (a) A fast wavelet transform is computed with a cascade of filterings with $\overline{h}$ and $\overline{g}$ followed by a factor 2 sub-sampling; (b) A fast inverse wavelet transform reconstructs progressively each a_j by inserting zeros between samples of a_{j+1} and d_{j+1}, filtering and adding both sequences.

This calculation is a straightforward iteration of the DWT algorithm described above. At the root of the tree we initialize $c_0^0[n] = \hat{x}[n]$ with the signal approximation in Eq. 3.26. Then, at the decomposition we have

$$c_{j+1}^{2k}[l] = \boldsymbol{c}_j^k * \overline{\boldsymbol{h}}_0[2l] \quad \text{and} \quad c_{j+1}^{2k+1}[l] = \mathbf{c}_j^k * \overline{\boldsymbol{h}}_1[2l], \tag{3.35}$$

whereas at the reconstruction

$$c_j^k[l] = \check{\boldsymbol{c}}_{j+1}^{2k} * \boldsymbol{g}_0[l] + \check{\boldsymbol{c}}_{j+1}^{2k+1} * \boldsymbol{g}_1[l] \tag{3.36}$$

for $j = 0, \ldots J$; $k = 0, \ldots, 2^j - 1$ and $l = N2^{-j}$.

As well as DWT, the coefficients $\boldsymbol{c}_{j+1}^{2k}$ and $\boldsymbol{c}_{j+1}^{2k+1}$ are obtained by sub-sampling and convolution. Iterating these equations along the branches of a wavelet packet tree computes all wavelet packet coefficients (Figure 3.8). In a full wavelet-packet tree of depth $J = \log_2 N$ there are a total of NJ coefficients, so information is clearly redundant at each scale.

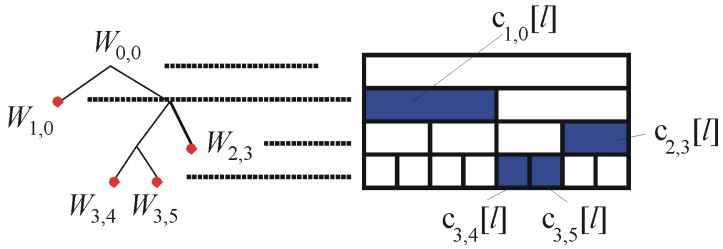


Figure 3.9: Example of an admissible tree for a sequence of N=8, decomposed up to a maximum scale of J=3. The selection of nodes is carried out in a way such that the full time and frequency axis is covered with N basis functions of different time-scale resolutions.

This tree constitutes a dictionary of orthonormal bases. A *dictionary of orthonormal bases* $\mathfrak{D}$ for $\mathbb{R}^N$ is a binary tree if it satisfies the following conditions:

- Each basis $c_j^k[l] \in W_{j,k}$ can be spanned within $I = [0, N)$ in the space (or time) domain, with subintervals $I_{j,k}$ of the form: $I_{j,k} = [2^{n_0-j}k, 2^{n_0-j}(k+1))$, for $j = 0, \ldots J$, $k = 0, \ldots, 2^j - 1$ and $J \leq n_0$.
- Assuming that $W_{j,k}$ is also represented by $I_{j,k}$ then $W_{j,k} = W_{j+1,2k} \oplus W_{j+1,2k+1}$.

In this scheme, there may be several node configurations choices that allow the perfect reconstruction of $\hat{x}[n]$, as long as the selection of nodes covers the full horizontal line, without any of them overlapped in the vertical line (Fig. 3.9). Each possible configuration of selected nodes leading to perfect reconstruction is referred to as "admissible tree", and requires of $O(N)$ computational resources. The Local Discriminant Bases algorithm (subsection 3.3.3) attempts to find an optimum admissible tree for classifying signals.

3.2.4 Block Transforms

Wavelet-packet bases are designed by dividing the frequency axis in intervals of varying sizes. Thus, these bases are particularly adapted to decomposing signals that have different behavior in frequency intervals. However, if signals vary in time, it is then more appropriate to decompose them in a *block basis* that segments the time axis in window sizes adapted to the signal structures. This subsection explains how to generate block basis of $\mathbf{L}^2(\mathbb{R})$. The cosine basis described in subsection 3.2.4.3 defines particularly interesting block bases.

3.2.4.1 Block Orthogonal Basis

Discrete block bases are obtained by dividing the time axis in consecutive intervals $[a_p, a_{p+1} - 1]$ by means of rectangular windows of the form

$$g_p[n] = \mathbf{1}_{[a_p, a_{p+1}-1]}(n), \tag{3.37}$$

for all $p \in \mathbb{Z}$ and $a_p \in \mathbb{Z}$. For signals of dyadic length N, a common choice is to use p intervals of width $l = N2^{-j}$. In this case, the window boundaries are set to $a_p = pl$ for $0 \leq p \leq 2^j$.

Assuming that $\{e_{k,l}\}_{0 \leq k < l}$ is an orthonormal basis of $\mathbb{C}^l$ for any $l > 0$, it can be shown that the set

$$\{g_{p,k}[n] = g_p[n]\, e_{k,l}[n - a_p]\}_{0 \leq k < l, p \in \mathbb{Z}} \tag{3.38}$$

forms a block orthonormal basis on the space W_p. One set of discrete block basis is obtained with the Fourier bases

$$\left\{ e_{k,l}[n] = \frac{1}{\sqrt{l}} \exp\left(i \frac{2\pi k n}{l} \right) \right\}_{0 \leq k < l}. \tag{3.39}$$

The block transform for this basis then becomes

$$X[p,k] = \sum_{n=0}^{N-1} x[n]\, g^*_{p,k}[n]. \tag{3.40}$$

This transform is a variant of the STFT with the difference that the range of admissible frequencies in the basis functions are restricted to the window size. This characteristic enables perfect reconstruction of $x[n]$ which is obtained as

$$x[n] = \sum_p \sum_k X[p,k]\, g_{p,k}[n]. \tag{3.41}$$

The resulting block of Fourier vectors $g_{p,k}$ generate sharp transitions at the window border that alter the frequency localization, producing generally large coefficient values in this zone.

3.2.4.2 Cosine and Sine Transforms

Similar to the DFT, discrete cosine and sine transforms (DCT/DST) also have the ability to express transient signal components in terms of sinusoids of different frequencies and magnitudes. The obvious distinction is that the former only uses real-valued cosine and sine functions, instead of complex exponentials.

Discretizing the frequency domain of DCT/DST can also be thought of as a way of implicitly extending the original signal beyond the last sample point. However, while DFT doubles the length assuming periodical extensions of $x[n]$, in the DCT/DST this extension can be either symmetric or asymmetric in relation to both ends of the signal.

For this reason, two issues arise at the time of setting the DCT/DST: 1) the type of extension at both the left and right end of the signal; and 2) the sample position this extension is applied. Extensions can be symmetric or asymmetric at both left and right ends of the signal, whereas the position can be around the last point or half way between two data points of both boundaries. Every combination corresponds to a specific type of DCT or DST. There is a total of sixteen possibilities, half of which correspond to eight DCTs (those where the *left* boundary is symmetric) and the other half to 8 DSTs (the left boundary is asymmetric).

However, not all of them have become practical in real applications. In particular, it is well known that processing a signal by pieces introduces coefficient distortion so that more coefficient terms are necessary to represent features accurately. This principle determines the DFT-based strategy for implementing the DCT/DST algorithm, which tries to reduce this distortion by some periodical extension of the initial signal. For all these reasons, only four DCT and four DST transforms (denoted as DCT-I/IV and DST-I/IV) are of practical utility, and constitute methodologies of this type of transform.

More formally, DCTs and DSTs implement invertible functions that transform a sequence of N real numbers $f : \mathbb{R}^N \to \mathbb{R}^N$. The four DCTs are

$$\text{DCT-I:} \quad X_I[k] = \frac{1}{2}\left\{x[0] + (-1)^k x[N-1]\right\} + \sum_{n=1}^{N-2} x[n] \cos\left(\frac{\pi n k}{N-1}\right) \quad k = 0, \ldots, N-1, \tag{3.42}$$

$$\text{DCT-II} \quad X_{II}[k] = \sum_{n=0}^{N-1} x[n] \cos\left[\frac{\pi k}{N}\left(n + \frac{1}{2}\right)\right] \quad k = 0, \ldots, N-1, \tag{3.43}$$

$$\text{DCT-III} \quad X_{III}[k] = \frac{1}{2}x[0] + \sum_{n=1}^{N-1} x[n] \cos\left[\frac{\pi n}{N}\left(k + \frac{1}{2}\right)\right] \quad k = 0, \ldots, N-1, \tag{3.44}$$

$$\text{DCT-IV} \quad X_{IV}[k] = \sum_{n=0}^{N-1} x[n] \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)\left(k + \frac{1}{2}\right)\right] \quad k = 0, \ldots, N-1, \tag{3.45}$$

and the four DSTs are

$$\text{DST-I:} \quad X_I[k] = \sum_{n=0}^{N-1} x[n] \sin\left[\frac{\pi}{N+1}(n+1)(k+1)\right] \quad k = 0, \ldots, N-1, \tag{3.46}$$

$$\text{DST-II:} \quad X_{II}[k] = \sum_{n=0}^{N-1} x[n] \sin\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)(k+1)\right] \quad k = 0, \ldots, N-1, \tag{3.47}$$

$$\text{DST-III:} \quad X_{III}[k] = \frac{(-1)^k}{2} x[N-1] + \sum_{n=0}^{N-2} x[n] \sin\left[\frac{\pi}{N}(n+1)\left(k+\frac{1}{2}\right)\right] \quad k = 0, \ldots, N-1, \tag{3.48}$$

$$\text{DST-IV:} \quad X_{IV}[k] = \sum_{n=0}^{N-1} x[n] \sin\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\right] \quad k = 0, \ldots, N-1. \tag{3.49}$$

In general, coefficient distortion is minimized when extensions are even (Fig. 3.10a). This explains why DCTs, and in particular the DCT-II generally have better coding performance than DFT and DSTs, yielding smaller amplitude coefficients at high frequency.

The use of DCT-II is widespread in image JPEG and video MPEG compression (Rao and Yip 1990). It is also equivalent (up to an overall scale factor of 2) to a DFT of $4N$ real inputs of even symmetry where the even-indexed elements are zero. That is, half of the DFT coefficients of the $4N$ inputs where $x[n] = 0$, $x[2n\text{-}1] = x[n]$ for $0 \leq n < N$, and $x[4N\text{-}n] = x[n]$ for $0 < n < 2N$. The inverse of DCT-II is DCT-III followed by a factor of $2/N$, but some authors prefer multiplying the term $X[0]$ by $1/\sqrt{2}$ and $X[k]$ by $\sqrt{2/N}$ for $1 \leq k \leq N-1$ in order to make DCT-II orthonormal. DCT-II then becomes

$$X_{II}[k] = \lambda_k \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x[n] \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right], \quad \text{with} \quad \lambda_k = \begin{cases} 2^{-1/2} & \text{if } k = 0 \\ 1 & \text{otherwise} \end{cases} \tag{3.50}$$

and is its inverse is

$$x[n] = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} \lambda_k X_{II}[k] \cos\left[\frac{k\pi}{N}\left(n+\frac{1}{2}\right)\right]. \tag{3.51}$$

In this case, DCT-II and DCT-III are equivalent, but the correspondence with DFT calculations is broken. Similarly, a factor of $\sqrt{2/N}$ produces orthonormal basis for the DCT-IV

$$X_{IV}[k] = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x[n] \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\right] \tag{3.52}$$

where

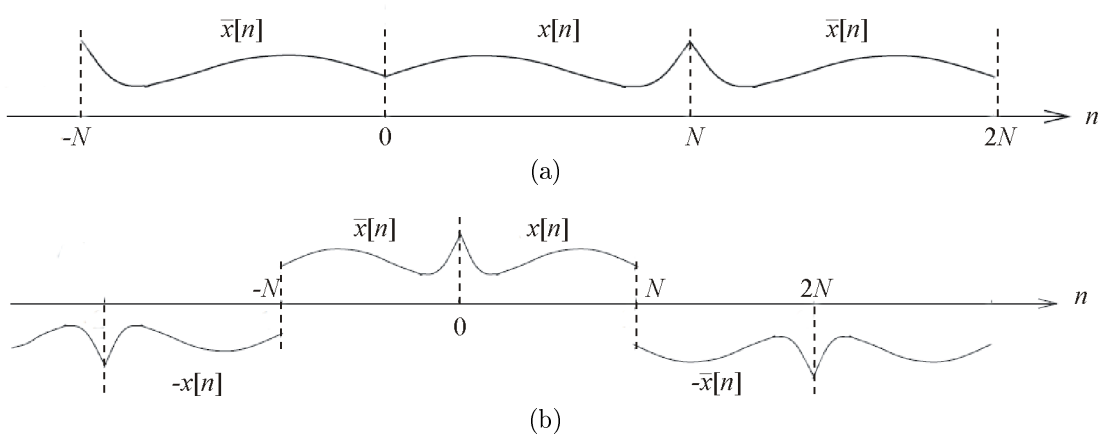


Figure 3.10: Signal extensions of the main cosinus transforms (from Mallat 1990); a) Extension for the DCT-II; b) Extension for the DCT-IV. A cosine II extends $x[n]$ into a signal of period 2 which is symmetric around 0. The extension in a cosinus-IV, however, is asymmetric with respect to N and of period 4.

$$x\left[n\right] = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} X_{IV}\left[k\right] \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\right]. \tag{3.53}$$

DCT-IV is also the basis of the MP3 codec system (also known as *Modified Discrete Cosinus Transform,* MDCT), as extending asymmetrically at the right boundary (Fig. 3.10b) has demonstrated better compression performance in audio signals (Rao and Yip 1990).

3.2.4.3 Block Cosine Transform

With the previous definition of the DCT, deriving a block transform for cosine bases is straightforward. If the cosinus term is taken from 3.50 as the basis of 3.39, it leads to the corresponding block basis of DCT-II

$$\left\{ g_{p,k}\left[n\right] = g_p\left[n\right] \lambda_k \sqrt{\frac{2}{l}} \cos\left[\frac{k\pi}{l}\left(n+\frac{1}{2}-a_p\right)\right] \right\}_{0 \le k < N \in \mathbb{Z}}. \tag{3.54}$$

This is the only cosine function that is practical in a block transform. The block cosine-IV, on the other hand, would produce very large coefficients with slow decay, as the extended signal x normally present discontinuities beyond the interval. The next section explains how to introduce smooth windows to circumvent this problem.

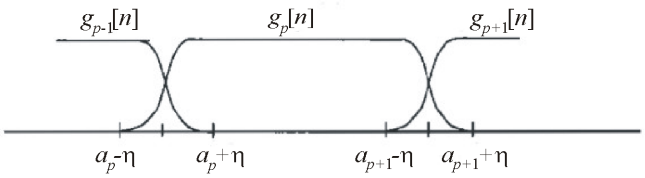


Figure 3.11: Each window g_p has a support $[a_p - \eta, a_{p+1} + \eta]$ with an increasing profile and decreasing profile over $[a_p - \eta, a_p + \eta]$ and $[a_{p+1} - \eta, a_{p+1} + \eta]$.

3.2.4.4 Lapped Projectors

Lapped projectors are implemented by introducing windows $g_p[n]$ in the basis functions which overlap at the time interval $[a_p - \eta, a_p + \eta]$. Let us divide the interval I_p in two overlapping intervals O_p, O_{p+1} and a central interval C_p:

$$I_p = [a_p - \eta, a_{p+1} + \eta] = O_p \cup C_p \cup O_{p+1} \tag{3.55}$$

with

$$O_p = [a_p - \eta, a_{p+1} + \eta] \quad \text{and} \quad C_p = [a_p + \eta, a_{p+1} - \eta]\,. \tag{3.56}$$

The space W_p is characterized by a window g_p whose support is I_p and which has a raising profile on O_p and a decaying profile on O_{p+1}:

$$g_p[n] = \begin{cases} 0 & \text{if } n \notin I_p \\ \beta\left[\eta^{-1}(n - a_p)\right] & \text{if } n \in O_p \\ 1 & \text{if } n \in C_p \\ \beta\left[\eta^{-1}(a_{p+1} - n)\right] & \text{if } n \in O_{p+1} \end{cases} \tag{3.57}$$

where the profile β can be any monotone increasing function of the form

$$\beta[u] = \begin{cases} 0 & \text{if } u < -A \\ 1 & \text{if } u > A \end{cases} \tag{3.58}$$

such that

$$\beta^2[u] + \beta^2[-u] = 1 \quad \forall u \in [-A, A] \tag{3.59}$$

This window is illustrated in Figure 3.11. An example of the β profile can be the function

$$\beta_0[n] = \sin\left(\frac{\pi}{4}\left(1 - \frac{u}{A}\right)\right) \quad \forall u \in [-A, A], \tag{3.60}$$

but its derivative at $u = \pm A$ is non-zero so β is not differentiable at $\pm A$ and still introduces some coefficient distortion. Windows of better regularity are constructed with a profile β_r defined by induction for $r \geq 0$ by

$$\beta_{r+1}[u] = \beta_r[u] \sin\left(\frac{\pi u}{2A}\right) \quad \forall u \in [-A, A]. \tag{3.61}$$

For any $k \geq 0$, one can verify that β_r satisfies Eq. 3.59 and has $2^r - 1$ vanishing derivatives at $u = \pm A$. The resulting β, and therefore g_p, are $2^r - 1$ times continuously differentiable.

3.2.4.5 Local Cosine Basis

Lapped orthogonal bases are block basis constructed from smooth windows $g_p[n]$ such as those defined in 3.57. To avoid coefficient distortion, the basis vectors are extended over n with a symmetry which depends on the shape of $\{e_{k,l}[n]\}_{0 \leq k < l}$.

In this sense, the discrete cosine-IV bases

$$\left\{ e_{k,l}[n] = \sqrt{\frac{2}{l}} \cos\left[\frac{\pi}{l}\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)\right] \right\}_{0 \leq k < l}. \tag{3.62}$$

have the advantage of including vectors that have a natural symmetric and antisymmetric extension, with respect to -1/2 and N-1/2, which extend any signal periodically. This produces the discrete local cosine basis of $\mathbf{l}^2(\mathbb{Z})$

$$\left\{ g_{p,k}[n] = g_p[n] \sqrt{\frac{2}{l}} \cos\left[\pi\left(k + \frac{1}{2}\right)\frac{n + 1/2 - a_p}{l}\right] \right\}_{0 \leq k < l, p \in \mathbb{Z}}, \tag{3.63}$$

which can be used in both 3.40 and 3.41 for signal decomposition and reconstruction, respectively.

Other cosine and sine functions may need different extensions in order to improve the right border treatment. In such cases, a common choice is to use DCT-II bases and extend either the input signal $x[n]$ or the basis $e_{k,l}$ as

$$\tilde{f}[n] = \begin{cases} f[n] & \text{if } n \in [0, N-1] \\ f[-1-n] & \text{if } n \in [-N, -1] \\ -f[2N-1-n] & \text{if } n \in [N, 2N-1] \\ -f[2N+n] & \text{if } n \in [-2N, -N-1] \end{cases}. \tag{3.64}$$

3.2.4.6 DCPT. Cosinus Packet Transform Computation

The previous framework led to the definition of local cosine functions by Coifman and Meyer (1991), who demonstrated that restricting the time interval to dyadic sizes allows tree-structured cosinus-based transforms to be constructed in a similar way to those of DWPT. Therefore, local cosine bases can be adaptively chosen with the fast dynamical selection algorithm that will be described in subsection 3.3.3.

Let $a_{j,k}$ denote subdivision points located at half integers

$$a_{j,k} = kN2^{-j} - 1/2 \quad \text{for } 0 \leq k \leq 2^j. \tag{3.65}$$

For any $j \geq 0$, the interval [0,N] is covered by 2^j windows with support $[a_{k,j} - \eta, a_{k+1,j} + \eta]$:

$$g_{k,j}[n] = \begin{cases} \beta\left[\eta^{-1}\left(n - a_{j,k}\right)\right] & \text{if } n \in [a_{j,k} - \eta, a_{j,k} + \eta] \\ 1 & \text{if } n \in [a_{j,k} + \eta, a_{j,k+1} - \eta] \\ \beta\left[\eta^{-1}\left(a_{j,k+1} - n\right)\right] & \text{if } n \in [a_{j,k+1} - \eta, a_{j,k+1} + \eta] \\ 0 & \text{otherwise} \end{cases} \tag{3.66}$$

To ensure that the support of $g_{k,j}$ is included in [0,N] for k=0 and $k = 2^j - 1$, the left and right sides of these windows are modified by setting $g_{0,j} = 1$ if $n \in [0, \eta]$, and $g_{2^j-1,j} = 1$ if $n \in [N - \eta, N]$. The size of the raising and decaying profile η in $g_{k,j}$ is independent of j. However, to guarantee that windows overlap only with their two neighbors, the length $l = a_{j,k+1} - a_{j,k} = N2^{-j}$ must be greater than the size 2η of the overlapping intervals and hence

$$\eta \leq N2^{-j-1}. \tag{3.67}$$

Similar to wavelet-packet trees, a local cosine tree is constructed by recursively dividing spaces built with local cosine bases. As such, a tree node at depth j and position k is associated to a vector space $W_{j,k}$ generated by the local cosine family

$$W_{j,k} = \left\{ h_{j,k,l}[n] = g_{k,j}[n] \sqrt{\frac{2}{N2^{-j}}} \cos\left[\pi \left(l + \frac{1}{2}\right) \frac{n - a_{k,j}}{N2^{-j}}\right] \right\}_{0 \leq l < N2^{-j}}, \tag{3.68}$$

where k and l now exchange frequency and position index, with respect to DWPT. The discrete cosine packet transform (DCPT) is readily obtained as

$$\mathrm{DCPT}\left\{\hat{x}[n]; 2^j, k2^j\right\} = c_{j,k}[l] = \sum_{n=0}^{N-1} x[n]\, h_{j,k,l}[n] \tag{3.69}$$

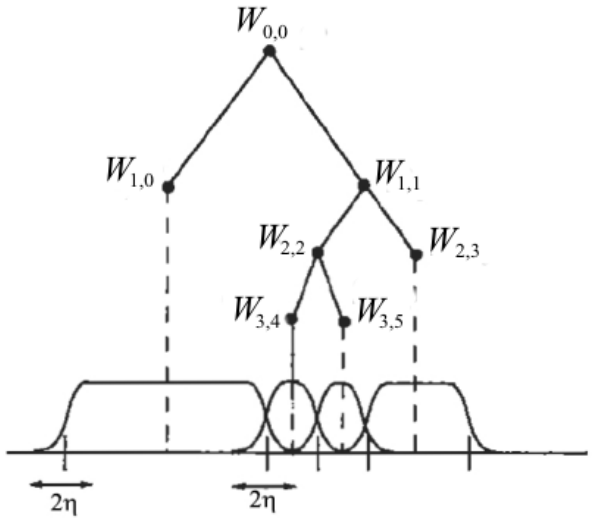


Figure 3.12: An admissible binary tree of local cosine spaces divides the time axis in windows of dyadic lengths.

It can be shown that any $x \in W_{j,k}$ has support in $[a_{j,k} - \eta, a_{j,k+1} + \eta]$, that $W_{j,k}$ is orthogonal to $W_{j,q}$ for $k \neq q$ and that

$$W_{j,k} = W_{j+1,2k} \oplus W_{j+1,2k+1}. \tag{3.70}$$

This splitting property implies that the union of local cosine families $W_{j,k}$ at the nodes of any admissible sub-tree is an orthonormal basis of $W_{0,0} = \mathbb{C}^N$ (Mallat 1990). However, in the DCPT this union is obtained from time divisions of various lengths, as illustrated in Fig 3.12. The signal $\boldsymbol{x}$ can then be recovered from the local cosine coefficients at the leaves of the admissible tree as

$$x\,[n] = \sum_j \sum_k \sum_l c_{j,k}\,[l]\,h_{j,k,l}\,[n]\,. \tag{3.71}$$

The windows $\boldsymbol{g}_{j,k}$ of local cosine trees have raising and decaying profile of equal size at all scales 2^j. If η is small compared to the interval size $N2^{-j}$ then $\boldsymbol{g}_{j,k}$ has a relatively sharp variation at its border that can create large coefficients up to a frequency order of π/η, even though $\boldsymbol{x}$ may be smooth over $[a_{j,k}, a_{j,k+1}]$. To reduce the quantity of large coefficients we must increase η, but this also increases the minimum size of the window required, by $N2^{-j} = 2\eta$. The choice of η is therefore a trade-off between window regularity and maximum time resolution. This limitation is not present in the construction of wavelet-packet bases.

On the other hand, the translation of $\boldsymbol{x}$ also modifies the coefficient values of trigonometric bases, making therefore this type of transform not translation invariant. In this case, the problem is that a translation of any cosine component spreads over vectors of different frequencies, questioning its use in pattern recognition applications.

3.3 On Feature Extraction Methods

In principle, all the description methods outlined above are not intended to evaluate signal structures. They modify the coordinate system to another axis of representation that can serve to reveal hidden structural components. However, determining whether these components are useful for solving the problem at hand requires further analysis.

The analysis of structure information can be addressed in two ways, either in terms of "signal similarities" (large components) or in terms of "signal singularities" (class differences). For example, *Principal Component Analysis* (PCA) uses the covariance matrix as a similarity measure to detect the best features for compressing information, whereas the measure used in *Linear Discriminant Analysis* (LDA), the Fisher's separability measure (Fisher 1936), is used to represent the class differences.

This subsection reviews some tools intended for developing data analysis. All of them have in common the use of cost functions (or statistical information measures) which are used to evaluate features and guide the selection process. When they take part in pattern recognition systems, the role of automatic feature extraction is provided, freeing thus the user from performing a manual selection of coefficients. Although PCA and LDA are two popular techniques used for this purpose, their main limitation lies in their lack of ability to explain physical phenomena in data. The *Best-Basis* paradigm, and specially the *Joint Best-Basis* and *Local Discriminant Basis* algorithms (LDB) are more recent designs that address this problem.

3.3.1 Best-Basis Selection (BB)

The Best-Basis algorithm (Coifman and Wickerhauser 1992) was developed mainly for signal compression. This method first expands a given single signal into a specified dictionary of orthonormal bases. Then a complete basis which minimizes the entropy cost function is searched in the wavelet packet binary tree using a divide-and-conquer algorithm.

More precisely, let $A_{j,k}$ denote a set of basis vectors belonging to space $W_{j,k}$ arranged as a matrix

$$A_{j,k} = \left(\boldsymbol{w}_{j,k,0}, \ldots, \boldsymbol{w}_{j,k,N2^{-j}-1}\right)^{\mathrm{T}}. \tag{3.72}$$

where T denotes matrix transpose. Here, $w_{j,k,l}[n]$ can be vectors from a wavelet packet dictionary $\psi_j^k[n - 2^j l]$ or local cosine/sine functions $h_{j,k,l}[n]$. Let $\jmath(\boldsymbol{p})$ be an information cost function measuring the "goodness" of compression of a sequence $\boldsymbol{p}$. A common choice is the *Shannon* entropy

$$\jmath(\boldsymbol{p}) = \sum_i - |p_i|^r \log_2 |p_i|^r. \tag{3.73}$$

where the convention $0 \log_2 0 = 0$ is used. (From now on, we use "log" for the logarithm of base 2) Often r =1 or $r = 2$ is used. In this thesis we always use $r = 2$.

Now, let $B_{j,k}$ denote the best basis for the signal $\boldsymbol{x}$ restricted to the span of $A_{j,k}$ and assume that the set $B = (\boldsymbol{w}_0, \ldots, \boldsymbol{w}_{N-1})^{\mathrm{T}} \in \mathbb{R}^{N \times N}$ is initialized to store N basis functions from $\{\boldsymbol{w}_{j,k,l}\}$ corresponding to an admissible tree. Here, we changed $\boldsymbol{w}_{j,k,l}$ to $\boldsymbol{w}_n$ for notation convenience. Then, the best-basis (BB) $\mathfrak{B}$ corresponds to the admissible tree from the full binary tree that minimizes the measure $\jmath(B\boldsymbol{x})$

$$\mathfrak{B} = \arg\min_{B \in \mathfrak{D}} \jmath(B\boldsymbol{x}). \tag{3.74}$$

where $B\boldsymbol{x}$ also denotes expansion coefficients for each $\boldsymbol{w}_n$ computed from 3.35. The following divide-and-conquer algorithm essentially performs 3.74 by comparing the efficiency of each parent node and its two children nodes:

Algorithm 3.1 Best-Basis (BB). *Given a vector* $\boldsymbol{x}$:

- **Step 0:** *Choose a dictionary of orthonormal bases* $\mathfrak{D}$ (*i.e. specify QMFs for a WPT dictionary or decide to use either a local cosine or sine dictionary*) *and specify the maximum depth of decomposition* J *and an information cost function* $\jmath$.
- **Step 1:** *Expand* $\boldsymbol{x}$ *into the dictionary* $\mathfrak{D}$ *and obtain the coefficients* $\{A_{j,k}\boldsymbol{x}\}_{0 \le j \le J, 0 \le k \le 2^j - 1}$.
- **Step 2:** *Set* $B_{J,k} = A_{J,k}$ *for* $k = 0, \ldots, 2^J - 1$.
- **Step 3:** *Determine the best subspace* $B_{j,k}$ *for* $j = J - 1, \ldots, 0$; $k = 0, \ldots, 2^j - 1$ *by*

$$B_{j,k} = \begin{cases} A_{j,k} & \text{if } \jmath(A_{j,k}\boldsymbol{x}) \le \jmath(B_{j+1,2k}\boldsymbol{x} \cup B_{j+1,2k+1}\boldsymbol{x}), \\ B_{j+1,2k} \cup B_{j+1,2k+1} & \text{otherwise.} \end{cases} \tag{3.75}$$

After this "pruning" process the best compressing features, which respond to a basis configuration of an admissible tree, are stored in $B_{0,0}$. In addition, since the measurement $\jmath$ is additive

$$\jmath\left(B_{j+1,2k}\boldsymbol{x} \cup B_{j+1,2k+1}\boldsymbol{x}\right) = \jmath\left(B_{j+1,2k}\boldsymbol{x}\right) + \jmath\left(B_{j+1,2k+1}\boldsymbol{x}\right), \tag{3.76}$$

calculating measurements for joint nodes requires an addition at subsequent dyadic scales. Thanks to this property, the divide-and-conquer algorithm can be implemented rapidly.

3.3.1.1 Dimensionality Reduction

Although $\mathfrak{B}$ contains the coordinates with the highest possible compression efficiency, in practical applications it is preferable to use only the best k ($\leq N$) coordinates. This can be done by evaluating the dictionary as

$$\jmath\left(\mathfrak{B}\boldsymbol{x}\right) = \sum_{n=1}^{k} \jmath_{(n)} \tag{3.77}$$

where $\jmath_{(i)}$ denotes increasing rearrangement of $\jmath$. Another possibility is to take into account the components that carry the signal energy larger than a certain threshold $\rho > 0$:

$$\jmath\left(\mathfrak{B}\boldsymbol{x}\right) = \sum_{n=1}^{k} 1\left(\left|\boldsymbol{w}_{(n)}^{\mathrm{T}}\boldsymbol{x}\right|^2 \geq \rho\right) \jmath_{(n)}, \tag{3.78}$$

The automatic selection of features should be completed by using an algorithm that determines the optimal value for k. This problem turns out to be far more complicated in classification applications and, very recently it has been addressed by means of the so-called *minimum description length* principle (MDL) (Rissanen 1983, Saito and Woei 2005). Although this topic is beyond the scope this thesis and k has been manually fixed in all the experiments. This issue will be discussed in the final chapter.

3.3.1.2 Best-Basis Selection from a Library of Orthonormal Bases

Given a method for choosing the Best-Basis from a dictionary, an obvious question arises. How can we select the best dictionary from a library? The strategy of Coifman and Majid (1993) is very simple: pick the one giving the minimum entropy among them.

More precisely, let $\mathfrak{L} = \{\mathfrak{D}_1, \ldots, \mathfrak{D}_M\}$ denote a library of orthonormal bases where $\mathfrak{D}_m$ represents a dictionary of orthonormal bases. For each dictionary the best basis $\mathfrak{B}_m$ of the signal $\boldsymbol{x}$ is computed by Eq. 3.74. This generates M different sets of the expansion coefficients $\{\mathfrak{B}_m\boldsymbol{x}\}_{m=1}^{M}$. For each coefficient set, the entropy defined in Eq. 3.73 is computed and then

the dictionary which gives the minimum entropy among M entropy values ,$\mathfrak{D}^*$, is selected as the best dictionary. That is

$$\mathfrak{D}^* = \arg \min_{\mathfrak{D}_m \in \mathfrak{L}} \jmath(\mathfrak{B}_m \boldsymbol{x}). \tag{3.79}$$

3.3.1.3 Karhunen-Loève Basis and Joint Best-Basis

To compress a given set of signals $\{\boldsymbol{x}_i\}_{i=1}^M$ rather than a single signal, one of the well-known traditional methods is the *Karhunen-Loève transform* (Ahmed and Rao 1975, Fukunaga 1990). Let X be the input data matrix $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M) \in \mathbb{R}^{N \times M}$. Then, the Karhunen-Loève basis (KLB) is defined as the eigenvectors $\{\boldsymbol{\varphi}_i\} \in \mathbb{R}^N$ of the symmetric positive definite matrix called *sample autocorrelation matrix*

$$R_X \triangleq \frac{1}{M} \sum_{i=1}^{M} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} = \frac{1}{M} X X^{\mathrm{T}}. \tag{3.80}$$

This basis is widely known for presenting the minimum entropy among all the orthonormal bases associated with orthogonal transformations in $\mathbb{R}^N$ (Watanabe 1967). The efficiency of this coordinate system is measured by the entropy of the total energy distribution $\boldsymbol{\gamma}_X$, which is defined by the normalized diagonal vector of R_X

$$\boldsymbol{\gamma}_X \triangleq \frac{\operatorname{diag}(R_X)}{\|\operatorname{diag}(R_X)\|}. \tag{3.81}$$

Watanabe (1967) proved that transforming the original signal space as $C = \varphi^{\mathrm{T}} X$, produces the minimum entropy $\jmath\left(\boldsymbol{\gamma}_{\varphi^{\mathrm{T}} X}\right)$. As such, φ^{T} is often referred to as the KLB. *Principal Component Analysis* (PCA) is very similar to KLB but the covariance matrix Σ_X is preferred rather than R_X. The dependence of this method on the eigenvalue system creates significant problems. They include the computational cost of diagonalizing R_X or the sensitivity to the alignment of the signals.

Despite the analytical character of KLB, this tool is unable to capture localized features. So Wickerhauser proposed extending the Best-Basis paradigm of section 3.3.1 to the Joint-Best-Basis algorithm. The idea is to evaluate the energy distribution at any coefficient, defined as in 3.74, but changing the terms $A_{j,k}\boldsymbol{x}$ by

$$A_{j,k}\boldsymbol{x} = \sum_{i=1}^{M} \left(\boldsymbol{w}_{j,k,l}^{\mathrm{T}} \boldsymbol{x}_i\right)^2_{0 \le l < N2^{-j}}, \tag{3.82}$$

so that the divide-and-conquer algorithm can find an admissible tree minimizing such distribution.

A joint Best-Basis as well as a KLB can also be computed after subtracting the mean $(1/M)\sum_{i=1}^{M}\boldsymbol{x}_i$ from each signal. The entropy criterion used in the Best-Basis algorithm is good for signal compression. However, it may not be necessarily good for other problems. Section 3.3.3 extends this algorithm for classification methods which are addressed in chapter 4.

3.3.2 Linear Discriminant Analysis (LDA)

Unlike PCA, the goal of LDA is to preserve as much of the class discriminatory information as possible when the axis is rotated, by minimizing the scatter of samples *within* each class and maximizing the scatter *between* classes, simultaneously.

This comes from Fisher's Linear separability measure (Fisher 1936) which tries to obtain large values when the elements of the same class are close together as well as separated from those of other classes. In this case, however, the resulting vector map is not necessarily an orthogonal matrix as in PCA.

To be more precise, let $\boldsymbol{\mu}_c = 1/N_c\sum_{i=1}^{N_c}\boldsymbol{x}_i$ be a mean vector of class c. Then the total mean vector is calculated as $\boldsymbol{\mu} = \sum_{c=1}^{C}\pi_c\boldsymbol{\mu}_c$, where π_c is the probability of class c (which can be set to N_c/N without the knowledge of true prior probability). The scatter of samples within each class can be measured by the *within-class covariance matrix* $\Sigma_w = \sum_{c=1}^{C}\pi_c\Sigma_c$, where Σ_c denotes *sample covariance matrix* of class c:

$$\Sigma_c = \frac{1}{N_c}\sum_{i=1}^{N_c}\left(\boldsymbol{x}_i^{(c)} - \boldsymbol{\mu}_c\right)\left(\boldsymbol{x}_i^{(c)} - \boldsymbol{\mu}_c\right)^{\mathrm{T}}.$$

The scatter of mean vectors around the total mean can be measured by the *between-class covariance matrix*, Σ_b, is calculated as

$$\Sigma_b = \sum_{c=1}^{C}\pi_c\left(\boldsymbol{\mu}_c - \boldsymbol{\mu}\right)\left(\boldsymbol{\mu}_c - \boldsymbol{\mu}\right)^{\mathrm{T}}.$$

One way to find out how much the classes are separated consists in maximizing the Fisher's Linear separability index (Fisher 1936)$\left(\boldsymbol{w}^{\mathrm{T}}\Sigma_w^{-1}\boldsymbol{w}\right)\left(\boldsymbol{w}^{\mathrm{T}}\Sigma_b\boldsymbol{w}\right)$. This requires solving the generalized eigenvalue problem, in this case given by,

$$\Sigma_w^{-1}\Sigma_b\boldsymbol{w} = \lambda\boldsymbol{w},$$

to obtain the eigenvalues and eigenvectors denoted as λ and $\boldsymbol{w}$, respectively. Finally, the feature vector $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i$ is computed for each signal i.

LDA calculates a total of C-1 basis vectors per signal in order to extract discriminatory information. It is an optimal strategy if signal classes obey to multivariate normal distributions with different mean vectors and equal covariance matrices (Fukunaga 1990, McLachlan 1992), but this condition is very hard to assume in practice. In addition, orthogonality is lost and the signal structures represented by the eigenvectors do not have an easy interpretation.

LDA can also be regarded as a classifier if the class of the closest mean vector is assigned to each sample $\boldsymbol{x}_i$. That is,

$$r(\boldsymbol{x}) = \arg\min_c \left\{ \left\| \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\mu}_c \right\|^2 \right\}. \tag{3.83}$$

This approach differs slightly from the Bayes-based LDA classifier in the sense that the mean class vectors act as nearest neighbors which assign classes to input samples (see section 4.3.2).

3.3.3 The Local Discriminant Basis (LDB)

The LDB algorithm (Saito 1994) attempts to find the most efficient coordinate system for classification by looking for the orthonormal basis functions that maximize the energy of class distributions in the time-frequency plane.

More precisely, let $\left\{ \boldsymbol{x}_i^{(c)} \right\}_{i=1}^{N_c}$ be a training data set consisting for $c = 1, \ldots, C$ classes with N_c signals in each class. The time-frequency energy distributions are estimated for every basis function and each class as

$$\Gamma_{j,k,l}^{(c)} = \frac{\sum_{i=1}^{N_c} \left(\boldsymbol{w}_{j,k,l}^{\mathrm{T}} \boldsymbol{x}_i^{(c)} \right)^2}{\sum_{i=1}^{N_c} \left\| \boldsymbol{x}_i^{(c)} \right\|^2}. \tag{3.84}$$

Here, the square of the expansion coefficients is accumulated and normalized by the global energy in class c. Then, one way for determining discrimination capacity at each position in the tree is by using discriminant measures of the form $\jmath(p, q)$, where p and q denote elements from two different classes.

There are different choices for $\jmath$ depending on the nature of data and the application problem at hand. Examples are the Hellinguer distance, the J-divergence (Kullback and Liebler 1951) or the symmetric relative entropy which is defined as,

$$\jmath(p, q) = p \log_r \frac{p}{q} + q \log_r \frac{q}{p}. \tag{3.85}$$

Normally, r=2 and the zero log convention is used. For many other possible choices the reader is referred to Baseville (1989). In order to handle multiple classes, it is better to take

$\binom{C}{2}$ pairwise combinations of $\jmath$. Therefore, the overall discrimination power measured at any element in the nodes and the wavelet tree can be computed as

$$\delta_{j,k,l} = \sum_{a=1}^{C-1} \sum_{b=a+1}^{C} \jmath\left(\Gamma_{j,k,l}^{(a)}, \Gamma_{j,k,l}^{(b)}\right) \tag{3.86}$$

In this case, feature selection is developed by finding a set of basis functions in the time-frequency dictionary that maximizes this measure. Let $B = \{\boldsymbol{w}_0, ..., \boldsymbol{w}_{N-1}\}$ denote the admissible tree represented by the basis vectors necessary to reconstruct the input domain exactly, just as in BB, where $B \in \mathfrak{D}$. Let δ_n denote discrimination power for a single coordinate $\boldsymbol{w}_n$ in this set.

DWPT connects a node at one level (the parent node) to two adjacent nodes at the subsequent level (the child nodes). Parent and children nodes correspond to two different representations of equivalent domains, so multiple admissible trees B exists. Thus, the best tree discriminating among classes, denoted by $\mathfrak{B}$, corresponds to the set B with maximum δ. That is,

$$\mathfrak{B} = \arg\max_{B \in \mathfrak{D}} \delta(B\boldsymbol{x}) \tag{3.87}$$

where $\delta(B\boldsymbol{x})$ is calculated as

$$\delta(B\boldsymbol{x}) = \sum_{n=1}^{2^{n_0}} \delta_n, \tag{3.88}$$

3.3.3.1 Divide-and-Conquer

Because of the tree characteristics of the wavelet-packet decomposition scheme, the implementation of Eq. 3.87 requires a divide-and-conquer algorithm of different performance compared to that of BB. In this case, the pruning process must be developed eliminating those nodes whose low $\delta_{j,k}$ do not make the difference among disjoint classes and, therefore, are not useful for classification.

Let $A_{j,k}$ denote a set of basis vectors at the space $W_{j,k}$. Let $B_{j,k}$ denote the LDB which we are after, restricted to the span of $A_{j,k}$. Also, let $\Delta_{j,k}$ be a work array containing discriminant measures $\delta_{j,k}$, where $\delta_{j,k} = \sum_{l=0}^{2^{n_0-j}-1} \delta_{j,k,l}$. Then, for the training signals $\left\{\boldsymbol{x}_i^{(c)}\right\}_{i=1}^{N_c}$ the following procedure maximizes the discriminant measure on the time-frequency energy distribution of classes:

Algorithm 3.2: Local Discriminant Basis (LDB)

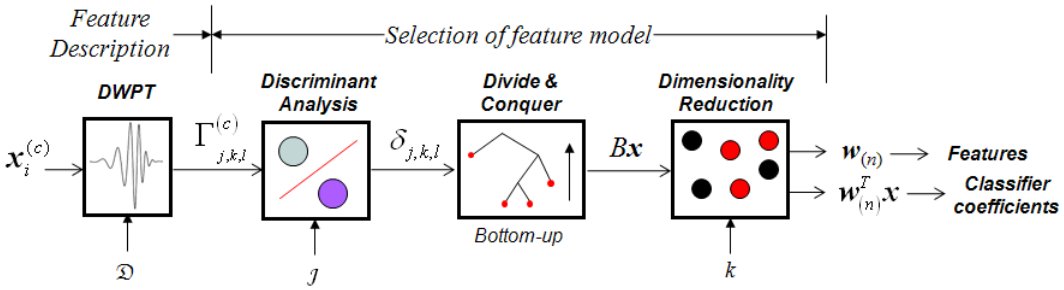


Figure 3.13: Flow diagram for the Local Discriminant Basis feature selection algorithm. The Best-Basis paradigm involves *description* and *selection* of features. Proposals for the three first blocks are considered in this thesis (see chapter 5).

- **Step 0:** *Choose a dictionary of orthonormal bases* $\mathfrak{D}$ (*i.e. specify* QMF's *for a wavelet packet dictionary or decide to use either the local cosine dictionary or the local sine dictionary*). *Specify the maximum level of decomposition* J *and a type of discriminant measure* $\jmath$.

- **Step 1:** *Construct time-frequency maps* Γ_c, *for* c = *1, ...,* C.

- **Step 2:** *Set* $B_{J,k} = A_{J,k}$ *and* $\Delta_{J,k} = \delta_{J,k}$ *for* $k = 0, \ldots, 2^J - 1$.

- **Step 3:** *Determine the best subspace* $B_{j,k}$ *for* $j = J-1, \ldots, 0$; $k = 0, \ldots, 2^j - 1$ *by the following rule:*

  - **Set** $\Delta_{j,k} = \delta_{j,k}$.
  - **If** $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$,
  - **then** $B_{j,k} = A_{j,k}$,
  - **else** $B_{j,k} = B_{j+1,2k} \cup B_{j+1,2k+1}$ and set $\Delta_{j,k} = \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$.

- **Step 4:** *Order the selected basis functions* $\boldsymbol{w}_n$ (*stored in* $B_{0,0}$) *by their power of discrimination.*

- **Step 5:** *Use* k ($\leq N$) *most discriminant basis functions for constructing classifiers.*

3.3.3.2 A Library of Local Discriminant Bases

Figure 3.13 shows the flow diagram of this algorithm. In this case, the dictionary is evaluated as

$$\delta\left(\mathfrak{B}\boldsymbol{x}\right) = \sum_{n=1}^{k} \delta_{(n)}, \tag{3.89}$$

and the values $\delta_{(i)}$ are arranged in decreasing order so as to ease the implementation of dimensionality reduction, but the selection of k is not evident since the classification performance decreases dramatically as basis functions of small δ are added to B.

Considering the basis of large energy components to attain certain degree of compression is a possibility, just as in the Best-Basis method. That is,

$$\delta\left(\mathfrak{B}\boldsymbol{x}\right) = \sum_{n=1}^{k} 1\left(\Gamma_{(n)} \geq \rho\right) \delta_{(n)}, \tag{3.90}$$

where $\Gamma_{(n)} = \frac{1}{\sum_{c=1}^{C} N_c} \sum_{i=1}^{M} \left| \boldsymbol{w}_{(n)}^{\mathrm{T}} \boldsymbol{x}_i \right|^2$ corresponds to the energy associated to $\boldsymbol{w}_{(n)}$ which has discrimination power $\delta_{(n)}$.

The preceding algorithm can be easily extended to a library of orthonormal bases. Let $\mathfrak{L} = \{\mathfrak{D}_1, \ldots, \mathfrak{D}_M\}$ denote a library, for $m = 1, 2, \ldots M$ dictionaries. Every LDB taken from $\mathfrak{D}_m$ has associated a discriminant value $\delta(\mathfrak{B}\boldsymbol{x})$ and . Let δ_m denote this value. Then, the optimal dictionary, $\mathfrak{D}^*$ (associated to the "best" LDB), is chosen by determining the index m which gives the maximum discriminant value. That is

$$\mathfrak{D}^* = \arg \max_{\mathfrak{D}_m \in \mathfrak{L}} \delta_m \tag{3.91}$$

3.4 Feature Representation Tools

When looking for patterns in data, it is sometimes useful to observe the physical structures associated to the coefficients, by means of a graphical tool. Two methods for the representation of time-frequency information and clusters, respectively, are described below.

The first method determines the time-frequency properties from the reconstructed basis functions and use them to indicate their support on the original signals. The basis functions are easily reconstructed by setting their corresponding coefficient cell as $c_{j,k}[l] = 1$. The second method is able to transform cluster data from a higher dimension, $\mathbb{R}^N$, to another dimension easily observed by the user (in 2D or 3D representation) so that he can have a better representation of the separability of classes from the higher domain.

3.4.1 Time-frequency Support

Wavelet packet as well as local cosine basis functions are indexed by means of three parameters (j, k, l) which are used as a reference to indicate their representation in the tree structure. Albeit rather roughly, it can be shown that these parameters are also a way of coding the signal properties that coefficients are representing. These properties are 1) the center of wiggles and main lobes of the wavelet shape (u), 2) the number of oscillations (ξ) and 3) both uncertainties of these parameters (σ_n and σ_ω, respectively), which serve to establish the extension support of the basis function in the time-frequency plane.

The momentum and variance of wavelet functions can be determined by two expressions equivalent to 3.9 and 3.11. That is,

$$u = \frac{1}{\left\|\boldsymbol{\psi}_{j,k,l}\right\|^2} \sum_{n=0}^{2^{n_0}} n \left|\psi_{j,k,l}\left[n\right]\right|^2 \tag{3.92}$$

and,

$$\sigma_n = \frac{1}{\left\|\boldsymbol{\psi}_{j,k,l}\right\|^2} \sum_{n=0}^{2^{n_0}} \left(n-u\right)^2 \left|\psi_{j,k,l}\left[n\right]\right|^2, \tag{3.93}$$

where , $\|\cdot\|$ denotes L^2 norm. Likewise, the frequency parameters, ξ and σ_ξ are obtained by using the frequency variable ω, instead of the position variable n, and the Fourier transform $\dot{\boldsymbol{\psi}}_{j,k,l}$ instead of $\boldsymbol{\psi}_{j,k,l}$.

These equations are crucial in the parametrization of signal structures and different artifacts. Since the triplet $\gamma = (j, k, l)$ indexes all the coefficients in the wavelet-packet tree, their associated basis functions $\psi_\gamma\left[n\right]$ can be reconstructed by setting the corresponding node cell $c_{j,k}\left[l\right]$ to one (with the remaining elements set to zero) and using the inverse DWPT (IDWT). Therefore, γ can also be seen as a way for parametrizing these structure patterns.

The index j denotes bandwidth resolution of the basis $\psi_\gamma\left[n\right]$ in the time-frequency plane, whereas k and l denote center frequency and center time position $(u = 2^j l)$, respectively. In general, the greater j is the better the frequency resolution (low σ_ξ), and vice-versa (the lower j is the better the time resolution). Since the successive down-sampling operations cause aliasing and alter the frequency content extracted with the filter sequences $\boldsymbol{h}$ and $\boldsymbol{g}$, the indexation of k must be corrected permuting it with the "Gray code". (Wickerhauser 1995).

3.4.2 Visualization of Data in $\mathbb{R}^N$

When evaluating the classification accuracy of features and/or contrasting their discriminant values, it is sometimes useful to display the provision of samples. One way to take a quick look at this provision is by plotting scatter maps but this form of interpretation becomes more difficult to observe and interpret above the 3rd dimension. A way for representing clusters of several dimensions in a 2D (or 3D) dimension preserving most of the important structure is provided by the Sammon maps (1969).

More precisely, assume that we have M vectors corresponding to input data in an N-space, $\boldsymbol{x}_i \in \mathbb{R}^N$ for $i = 1, ..., M$, and, corresponding to these, M output vectors in a D-space, $\boldsymbol{y}_i \in \mathbb{R}^D$, are defined. Here, D can be either 2 or 3 depending on a bi-dimensional or three-dimensional representation, respectively. The following non-linear mapping algorithm (NLM) minimizes an error ε that represents how well the points in $\mathbb{R}^D$ fit the original points in $\mathbb{R}^N$, and uses a steepest descent procedure to adjust $\boldsymbol{y}_i$ accordingly.

Let $d_{ij}^{(\boldsymbol{f})}$ be an Euclidean distance measure between two vectors $\left[\boldsymbol{f}_i, \boldsymbol{f}_j\right] \in \mathbb{R}^q$, defined as

$$d_{ij}^{(\boldsymbol{f})} = \sqrt{\sum_{k=1}^{q} \left(f_{ik} - f_{jk}\right)^2},$$

where f_{ik} is used to denote every single element from the i-th vector and its k-th dimension (here, $p = 1, \ldots, M$; and $q = N$ for the input $\boldsymbol{x}$ and $q = D$ for the output $\boldsymbol{y}$).

First, an initial guess for the outputs $\boldsymbol{y}_{pq}(0)$ in D-space must be provided. Although these vectors could be fixed randomly, it is preferable to project the inputs $\boldsymbol{x}$ orthogonally onto a D-space spanned by the D original coordinates of largest variance for a fast convergence of the algorithm. Then, inter-point distance measurements between $d_{ij}^{(\boldsymbol{x})}$ and $d_{ij}^{(\boldsymbol{y})}$ are compared to reduce ε and modify $\boldsymbol{y}_i$ accordingly in each iteration. Let $\varepsilon(t)$ define this error calculated after the t-th iteration

$$\varepsilon(t) = \frac{1}{c} \sum_{i \neq j}^{M} \frac{\left[d_{ij}^{(\boldsymbol{x})} - d_{ij}^{(\boldsymbol{y})}(t)\right]^2}{d_{ij}^{(\boldsymbol{x})}}. \tag{3.94}$$

where, $c = \sum_{i \neq j}^{M} d_{ij}^{(\boldsymbol{x})}$.

In order to modify y_{pq} at step t +1 and minimize 3.94, the output is recalculated as

$$y_{pq}(t+1) = y_{pq}(t) - \alpha \Delta_{pq}(t) \tag{3.95}$$

where

Object	Conventional concept	"Library" concept
Coordinate System	Standard Euclidean Basis (STD) Fourier Basis (FT,STFT)	Wavelet Packet Bases (WPT) Local Trigonometric Bases (CPT)
Compression	Karhunen-Loève Basis (PCA)	Joint Best-Basis (BB)
Classification	Linear Discriminant Analysis (LDA)	Local Discriminant Basis (LDB)

Table 3.1: Summary of the correspondences between the conventional concepts and the new concepts based on the Best-Basis paradigm (or "library" of bases) reviewed or discussed in this chapter.

$$\Delta_{pq}(t) = \frac{\partial \varepsilon(t)}{\partial y_{pq}(t)} \Big/ \left| \frac{\partial^2 \varepsilon(t)}{\partial y_{pq}(t)^2} \right| \tag{3.96}$$

and α is a factor (normally configured empirically to 0.3 or 0.4, depending on the desired convergence). The partial derivatives are given by

$$\frac{\partial \varepsilon}{\partial y_{pq}} = -\frac{2}{c} \sum_{\substack{j=1 \\ j \neq p}}^{N} \left[\frac{d_{pj}^{(\boldsymbol{x})} - d_{pj}^{(\boldsymbol{y})}}{d_{pj}^{(\boldsymbol{x})} d_{p,j}^{(\boldsymbol{y})}} \right] (y_{pq} - y_{jq}) \tag{3.97}$$

and

$$\frac{\partial^2 \varepsilon}{\partial y_{pq}^2} = -\frac{2}{c} \sum_{\substack{j=1 \\ j \neq p}}^{N} \frac{1}{d_{pj}^{(\boldsymbol{x})} d_{pj}^{(\boldsymbol{y})}} \left[\left(d_{pj}^{(\boldsymbol{x})} - d_{pj}^{(\boldsymbol{y})} \right) - \frac{(y_{pq} - y_{jq})^2}{d_{pj}^{(\boldsymbol{y})}} \left(1 + \frac{d_{pj}^{(\boldsymbol{x})} - d_{pj}^{(\boldsymbol{y})}}{d_{pj}^{(\boldsymbol{y})}} \right) \right]. \tag{3.98}$$

These calculations can be iterated up to a certain value of t, or until ε converges at some desired threshold. The error value $\varepsilon(T)$ at the end of the process can be a reference indicating the accuracy of the new representation. Although the Sammon method can also be regarded as a feature extraction method, given to the fact that the dimensionality of input data is also reduced, the minimization process is more related to the "goodness of fit" of the representation in the new domain. For this reason it is very used as a display tool rather than a pattern recognition technique.

3.5 Conclusions

In this chapter, several methods related to feature extraction using a library of bases have been reviewed. Table 3.1 summarizes the correspondences discussed in this chapter. The

philosophy, also referred to as the Best-Basis paradigm, consists in selecting the best coordinate system of representation from a dictionary of basis functions, which contains the words necessary to describe to describe the problem at hand. Wavelet-packet dictionaries (including wavelet bases) and trigonometric bases are the main constituents of this library.

These bases are able to capture local features in data up to a certain accuracy, and provide an interpretation of signal irregularities much closer to the physical structures. By combining them with discriminant analysis, their degree of similarity (or difference) in relation to different signal classes can also be evaluated. As such, they should provide better understanding and insight into underlying phenomena compared to other conventional techniques.

The unification of the statistical and the library-based syntactic approach is intended to solve difficult problems in image analysis and signal processing applications in the near future, as the library concept is vastly more efficient compared with strategies confined to a single basis. In addition, such strategy allow us to find small feature sets that reduce the computational problems of many classification methods as well as improving their accuracy.

Chapter 4

Classifiers

4.1 Introduction

Information can come in the form of different objects: statistical data, measurements, sensor signals or random variables; and the the classification problem consists in predicting which class each object is in.

This chapter looks at supervised classifiers. The word 'supervised' means that the object class is known and used for two operations: 1) the configuration of the algorithm; and 2) the estimation of accuracy in the detection of classes. The first stage is often referred to as the "learning stage" (or training) and depends on the framework of the chosen classification method. The second stage, however, is inherent to all supervised classifiers and estimates the efficiency that the configured classifier would obtain in a real experiment of unknown classes. For this reason, this operation is referred to as the "validation stage".

A more formal definition of a classifier is provided in the next section, where the estimation of accuracy will be the main topic of discussion. Traditionally, the Bayes probability concept has been the reference tool for building classifiers (Bayes 1764). The linear methods, such as the naive Bayes classifier, the LDA-based method or the k-Nearest Neighbor are some examples of these. They are described in section 4.3.

In contrast, the more "exotic" classifiers, such as Classification-and-Regression-Trees (CART) or Neural-Networks (NN), base their decision rules on very complex networks of nonlinear functions and will not be considered in this thesis. The *Learning Vector Quantization* method (Kohonen et al. 1992) is a special case, since its performance lies in the middle of both approaches, making it particularly interesting for the fish experiments of this work. This method is described in section 4.4.

4.2 Problem Formulation

Let X be the space of all possible measurement vectors $\boldsymbol{x} \in \mathbb{R}^T$ and assume that each measurement $\boldsymbol{x} \subset X$ can fall into one of C categories, say $(1, \ldots, C)$ classes. A systematic way of predicting class membership is a rule that assigns a class to every measurement vector. Thus,

Definition. 4.1a *A classifier* (or *mapping rule*) *is a function,* $r(\boldsymbol{x})$, *defined on* X *so that for every* $\boldsymbol{x} \subset X$, *a class* c *is assigned.* $r(\boldsymbol{x})=c$ for $c \in \{1, \ldots, C\}$.

$$r(\boldsymbol{x}) = c, \text{ for } c \in \{1, \ldots, C\} \tag{4.1}$$

This systematic definition differ from others in which classifiers are seen as disjoint subsets of the space X.

Definition. 4.1b *Let* $\mathcal{A}_c$ *define the subset from* X *in which* $\boldsymbol{x}$ *falls into category* c

$$\mathcal{A}_c = \{\boldsymbol{x};\ r(\boldsymbol{x}) = c\}, \tag{4.2}$$

where $\{\mathcal{A}_1, \ldots, \mathcal{A}_C\}$ *are disjoint and* $X = \cup\mathcal{A}_c$. *A classifier is a partition of* X *into* C d*isjoint subsets,* $\mathcal{A}_1, \ldots, \mathcal{A}_c$, $X = \cup\mathcal{A}_c$ *such that for every* $\boldsymbol{x} \subset \mathcal{A}_c$ *the predicted class is* c.

In principle, this framework is optimal for an idealistic precise estimation of accuracy. However, one of the main problems found when constructing real classifiers is that very often we are limited to reduced sets with little prospect of obtaining large classified samples. In the systematic case, the knowledge for constructing classifiers is based on past experience which is summarized by a *learning sample* containing the measurement vectors $\hat{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, where $\hat{X} \in X$, and its associated classification $(c_1, \ldots, c_N)$. That is,

$$\mathcal{L} = \{(\boldsymbol{x}_1, c_1), \ldots, (\boldsymbol{x}_N, c_N)\}, \tag{4.3}$$

where $\mathcal{L}$ denotes learning set.

4.2.1 Estimation of Accuracy

Given a classifier r, in order to estimate its accuracy $R(r)$ will be defined as the *true* misclassification rate. A more formal definition can be given to this parameter by looking for an answer to the question: What is truth and how can it be estimated? This value can be conceptualized in the following manner:

> *"Using a sample taken in an experiment from a real distribution, construct a classifier. Then, draw another sample from the same population as the original sample was drawn from, and use the classifier to predict classification by estimating the misclassified proportion"*

To make a more precise definition of the preceding concept a probability model is needed. Assume that a sample set $\mathcal{L}$ has been taken from the relevant population $X \times C$. Since this sample comes from the relevant population, the couple $(\boldsymbol{x}, c)$ allows the partitioning of $\mathcal{L}$ in different class disjoint subsets of the form $\mathcal{L} = \cup \mathcal{A}_j$. Let $P(\mathcal{A}, c)$ be a probability that an entity $\boldsymbol{x}$ taken at random from $\mathcal{L}$ belongs to class c. The interpretation of $P(\mathcal{A}, c)$ is that a case $\boldsymbol{x}$ drawn at random from the relevant population has probability $P(\mathcal{A}, c)$ that its measurement vector is in $\mathcal{A}$ and its associated class is c. The estimation of $P(\mathcal{A}, c)$ for each one of the classes serves to construct $r(\boldsymbol{x})$. Then, take another sample of measurement vectors $(\boldsymbol{x}_s, c_s)$, $\boldsymbol{x}_s \in X$, $c_s \in C$ from the relevant population $X \times C$, so as to ensure this new set is independent of $\mathcal{L}$. And finally, determine the misclassified proportion as

$$R(r) = P\left(r\left(\boldsymbol{x}_s\right) \neq c_s\right) \tag{4.4}$$

Such estimation of the misclassification error is optimal as long as the set $\mathcal{L}$ is representative of the true population in X. In practice, however, samples are habitually scarce and $\mathcal{L}$ is often used both to construct $r(x)$ and to estimate $R(r)$. Therefore, we use $\hat{R}(r)$ to refer to this internal estimate instead of $R(r)$, or the complement $1 - \hat{R}(r)$ to denote the proportion of correct classifications.

In general, three types of internal estimate are of particular interest for the evaluation of classification accuracy. The *re-substitution estimate* (Smith 1947) is commonly used as the first reference, but is the least accurate. This estimate uses all of the elements in $\mathcal{L}$ both to construct the classifier $r\left(\boldsymbol{x}\right)$ and to compute $\hat{R}\left(r\right)$. To put it into an equation form

$$\hat{R}(r) = \frac{1}{N} \sum_{i=1}^{N} 1\left(r\left(\boldsymbol{x}_i\right) \neq c_i\right) \tag{4.5}$$

where $1(\cdot)$ defines the indicator function to be one if the statement inside the parentheses is true, and zero otherwise.

The second estimate is known as the *hold-out* (or *split-sample*) method. Here, the cases in $\mathcal{L}$ are divided into two subsets: $\mathcal{L}_t$ (training set) and $\mathcal{L}_s$ (validation or testing set). Only the cases in $\mathcal{L}_t$ are used to construct r, whereas the elements in $\mathcal{L}_s$ serve to estimate $R(r)$. Assuming N_s cases in $\mathcal{L}_s$, the holdout estimate, $\hat{R}^{ho}(r)$, is calculated as

$$\hat{R}^{ho}(r) = \frac{1}{N_s} \sum_{i=1}^{N_s} 1\left(r\left(\boldsymbol{x}_i\right) \neq c_i\right) \tag{4.6}$$

In this method care needs to be taken so that the cases in $\mathcal{L}_s$ can be both independent of $\mathcal{L}_t$ and drawn from a representative distribution. A common way to ensure this is to take $\mathcal{L}_s$ as 1/3 of the cases in $\mathcal{L}$ at random, although there is no theoretical formalism to justify this 1/3, 2/3 split.

For smaller sample sizes, another method known as *K-fold cross-validation* is preferred (Stone 1974). Here, the cases in $\mathcal{L}$ are randomly divided into K subsets of as equal size as possible. Let $\mathcal{L}_k$ denote all these subsets, where $k = 1, ..., K$ and assume that a classifier for each set r_k is constructed using the sample set $\mathcal{L}$ – $\mathcal{L}_k$. This set contains the cases in $\mathcal{L}$ that are not included in $\mathcal{L}_k$. Since none of the cases in $\mathcal{L}_k$ are used in the construction of r_k, a holdout estimate can be obtained for each classifier by means of

$$\hat{R}^{cv}(r) = \frac{1}{K} \sum_{k=1}^{K} \hat{R}^{ho}(r_k) \tag{4.7}$$

where

$$\hat{R}^{ho}(r_k) = \frac{1}{N_k} \sum_{i=1}^{N_k} 1\left(r_k\left(\boldsymbol{x}_i\right) \neq c_i\right). \tag{4.8}$$

The K-fold cross-validation method is also known as the *leave-one-out* estimate (Lachenbruch 1965) when $K = N$. Here, only one element from $\mathcal{L}$ is set aside and the remaining N - 1 elements are used to construct r_k. Then using each n-th case in Eq. 4.7 allows an estimate $R(r)$ to be determined.

In general, since all classification methods attempt to minimize $\hat{R}(r)$, which is merely an estimation of the true $R(r)$, the true accuracy will differ from that of simulation results. In this sense, the re-substitution estimate is not practical because it produces very optimistic results. To understand this, assume that $r(\boldsymbol{x})$ is defined by the partitions, $\mathcal{A}_1, ..., \mathcal{A}_C$ constructed from all elements in $\mathcal{L}$ and the same elements are used to calculate $\hat{R}(r)$. Then, $\hat{R}(r) = 0$, but if new values of $\boldsymbol{x}$ different from $(\boldsymbol{x}_1, ..., \boldsymbol{x}_N)$ are assigned arbitrarily to one or the other $\mathcal{A}_c$, it is hard to believe that $R(r)$ is anywhere near zero.

K-fold cross-validation is widely accepted as a good estimator and occasionally substituted for the split-sample approach to reduce processing time when large data sets are available (Goutte 1997). Additionally, obtaining a mean value of $\hat{R}^{ho}(r)$ (or $\hat{R}^{cv}(r)$), instead of a single estimate, is preferable to avoid accuracy bias and variance. So, habitually, both

hold-out and k-fold cross-validation are repeated several times for different combinations of $\mathcal{L}_t$ and $\mathcal{L}_s$, averaging all results at the end of the process.

Another way to tackle the problem of bias in reduced data sets is to use the *bootstrap* method. In this case, the size of the original set $\mathcal{L}$ is increased deliberately by repeating some of the elements in $\mathcal{L}$ randomly, or creating new ones from changing some of its attributes whenever possible (Efron 1983). However, it has been reported that this operation may be very optimistic for certain classifiers (Breiman et al. 1984, Kohavi 1995).

In practice, most frequently other factors such as processing speed, time and cost of analysis influence decisions regarding this estimate. A summary and extensive bibliography concerning this topic can be found in Toussaint (1974).

4.2.2 The Bayes Rule

The major guide used in the construction of classifiers is the concept of the Bayes rule. This concept states that when data is drawn from a probability distribution $P(\mathcal{A}, c)$, then the form of the most accurate classifier can be given in terms of $P(\mathcal{A}, c)$. The rule that is inferred from this definition is called the *Bayes rule* and is denoted as $r_B(\boldsymbol{x})$.

More precisely, let us assume that a sample $(\boldsymbol{x},c)$, $\boldsymbol{x} \in X$, $c \in C$, is drawn from the probability distribution $P(\mathcal{A}, c)$ (we changed $\boldsymbol{x}_s$into $\boldsymbol{x}$ for notation convenience), i.e., $P(\boldsymbol{x} \in \mathcal{A}, c = C) = P(\mathcal{A}, c)$. Then, $r_B(\boldsymbol{x})$ will be a Bayes rule if for any other classifier $r(x)$,

$$P(r_B(\boldsymbol{x}) \neq c) \leq P(r(\boldsymbol{x}) \neq c). \tag{4.9}$$

and the probability value

$$R_B = P(r_B(\boldsymbol{x}) \neq c). \tag{4.10}$$

is known as the *Bayes misclassification rate.*

In general, the Bayes rule, $r_B(\boldsymbol{x})$, is obtained from the distribution $P(\mathcal{A}, c)$ assuming that prior class probabilities π_c are known. More precisely, if $\pi_c = P(c)$ for $c \in C$, the probability distribution of the cth class measurement vectors is of the form $P(\mathcal{A}|c) = P(\mathcal{A}, c)/\pi_c$ and is obtained from the density $f_c(\boldsymbol{x})$,

$$P(\mathcal{A}, c) = \int_{\mathcal{A}} f_c(\boldsymbol{x})\, d\boldsymbol{x}. \tag{4.11}$$

Then, for $\mathcal{A} \subset X$ the Bayes rule is defined as

$$r_B(\boldsymbol{x}) = c \text{ on } \mathcal{A}_c = \left\{\boldsymbol{x};\ f_c(\boldsymbol{x})\,\pi_c = \max_i (\boldsymbol{f}_i \pi_i)\right\}. \tag{4.12}$$

Under the assumption of (4.11), Breinman et al (1984) demonstrated that for any other classifier r,

$$P(r(\boldsymbol{x}) = c) \leq P(r_B(\boldsymbol{x}) = c) = \int \max_c [f_c(\boldsymbol{x})\,\pi_c]\, d\boldsymbol{x}, \tag{4.13}$$

Therefore, the Bayes misclassification rate is

$$R_B = 1 - \int \max_c (\boldsymbol{f}_c \pi_c)\, d\boldsymbol{x}. \tag{4.14}$$

The Bayes rule is also known as the maximum a-posteriori likelihood rule. That is, classify $\boldsymbol{x}$ as that c for which $f_c(\boldsymbol{x})\,\pi_c$ is maximum. Since R_B is the minimum misclassification rate attainable, knowing the class distribution is useful in order to generate simulated data and to compare the accuracy of different classifiers. However, neither π_c nor $f_c(\boldsymbol{x})$ are known in practice. The value π_c can be assumed as the proportion of class c cases in $\mathcal{L}$ or supplied through other knowledge of data, but finding $f_c(\boldsymbol{x})$ is considered the major problem in the development of classifiers. In the following sections we explain how some classification methods address this problem.

4.3 Bayes Classifiers

In general, a Bayes classifier is a single probabilistic classifier constructed from the Bayes theorem. Such a classifier, attempts to converge to the Bayes rule, using a very simple probabilistic model of $f_c(\boldsymbol{x})$ that is calculated from the learning sample $\mathcal{L}$. The most common way of doing this consists in taking $P(\mathcal{A}, c)$ as the density estimator of class c in the form of $P(\mathcal{A}, c) = P(\mathcal{A}|c)\,\pi_c$.

4.3.1 Joint and Naive Classifiers

Let $X = \{\boldsymbol{x}_i, c_i\}$ denote the input data set, where each vector $\boldsymbol{x}_i \in \mathbb{R}^T$ is associated to class c_i, and $i = 1, \ldots, N$. Assume that each class, $c = 1, \ldots, C$, has a total of N_c samples, so that the proportion of class c is $\pi_c = N_c/N$. A Bayes classifier is constructed by using a learning set, $\mathcal{L}_t$, and calculating the probability $P(\{x(1), \ldots, x(T)\}|c)$ as the proportion of cases in class c that $(x(1), \ldots, x(T))$ is *repeated*. Then, to predict the unknown class of a new sample $\boldsymbol{q} = (x(1), \ldots, x(T))$, simply choose the label containing the most records of $\boldsymbol{q}$. That is,

$$r_B(\boldsymbol{q}) = \arg\max_c \left(P\left(q(1) = x(1), \ldots, q(T) = x(T) \,|c\right) \pi_c\right). \tag{4.15}$$

Very often, this rule is referred to as the Joint Bayes classifier. In order to supervise the process, $\boldsymbol{q}$ is taken from the testing set $\mathcal{L}_s$. Since its true class is known, the misclassification accuracy R_B can be estimated with some of the previous validation techniques exposed in section 4.2.1.

Originally, the Bayes rule 4.15 was designed for categorical variables. In this case, each attribute in $\boldsymbol{q}$ is limited to only a few different values (say two or three, depending on the application case). For real numbers, however, if the full vector does not match any of the training set, then $P(\boldsymbol{q}|c) = 0$ for all c and the assigned class is either unknown or chosen at random.

This situation can be avoided partially if, instead of $P\left(\{q(1), \ldots, q(T)\} \,|c\right)$, we use $\prod_{i=1}^{T} P\left(q(i) = x(i) \,|c\right)$. This last definition is known as the Naive Bayes classifier since it makes strong assumptions of independence between attributes. However, for a considerable number of attributes this product tends to underflow, so the preferred version of the Naive classifier introduces logs to the mapping rule such as

$$r_B(\boldsymbol{q}) = \arg\max_c \left(\ln \pi_c + \sum_{t=1}^{T} \ln P\left(q(t) = x(t) \,|c\right)\right). \tag{4.16}$$

Of course, the Joint Bayes classifier can also be implemented under real-valued inputs, but it is necessary to discretize the attributes of $\boldsymbol{x}$ beforehand so that the distribution $P(\boldsymbol{q}|\mathrm{c})$ can be calculated. Several choices are available depending on assumptions regarding the structure of data. Methods such as *fixed width interval*, *1R* or *ChiMerge* - among others; are included in many statical software packages (Bishop 1995).

4.3.2 LDA-Based Classifiers

Classifiers based on linear discriminant analysis determine $P(\boldsymbol{x}/c)$ on the assumption of Normal distribution of multivariate data, which is characterized by the common pooled covariance matrix $\Sigma = \sum_{c=1}^{C} \pi_c \Sigma_c$, and whose classes are separated by class mean vectors $\{\boldsymbol{\mu}_c\}$. As usual, this information is obtained by means of a training sample $\mathcal{L}_t$ in order to construct the classifier, whereas the remaining set $\mathcal{L}_s$ is used for validation. The Bayes rule

$$r_B(\boldsymbol{q}) = \arg\max_c \left(\ln \pi_c + \boldsymbol{\mu}_c^{\mathrm{T}} \Sigma^{-1} \boldsymbol{q} - \frac{1}{2} \boldsymbol{\mu}_c^{\mathrm{T}} \Sigma^{-1} \boldsymbol{\mu}_c\right), \tag{4.17}$$

determines the class of the objects of this set and the misclassification error R_B establishes its accuracy.

In general, this version of LDA is widely accepted and used without regard to lack of normality. It can not handle categorical variables since linear distribution are considered in the prediction of $\boldsymbol{x}$. However, when more than two classes come into play distributions of data are difficult to interpret.

4.3.3 k-Nearest Neighbor (k-NN)

Perhaps, the most straightforward classifier in machine learning is the *Nearest Neighbor* classifier (Fix and Hodges 1951). This method develops the classification procedure in a very simple and intuitive way: it determines the nearest neighbors to the item being classified and assigns the most represented class among them. The examples from the learning set must always be available in memory. Thus in this approach, learning means memorizing examples. For this reason, it is sometimes referred as *Memory-based classification.*

As usual, we have an input set $\mathcal{L}$ of labeled samples, $\boldsymbol{x}_i \in \mathbb{R}^T$ and class $c_i \in C$, which is divided into a training set $\mathcal{L}_t$ and a testing set $\mathcal{L}_s$. The goal is to classify an example $\boldsymbol{q}$, in $\mathcal{L}_s$. Therefore, for each $\boldsymbol{x}_i \in \mathcal{L}_t$ we calculate the distances

$$d\left(\boldsymbol{q}, \boldsymbol{x}_i\right) = \sum_{t \in T} \delta\left(q\left[t\right], x\left[t\right]\right)_i \tag{4.18}$$

where $\delta\left(\boldsymbol{q}, \boldsymbol{x}\right)$ define some metric that serves to select the k nearest neighbors. There are a large range of possibilities for distance metrics that can be used in many classification methods. A basic version for continuous and discrete attributes would be:

$$\delta\left(q, x_i\right) = \begin{cases} 0, & t \text{ discrete and } u = x_i \\ 1, & t \text{ discrete and } u \neq x_i \\ \left|u - x_i\right|, & t \text{ continuous} \end{cases} \tag{4.19}$$

This is a special case of the Minkowski Distance metric, also known as the Manhattan distance or the 1-norm (L_1). The general formula for such distance is

$$d_p\left(\boldsymbol{q}, \boldsymbol{x}_i\right) = \left(\sum_{t \in T} \left|q\left[t\right] - x_i\left[t\right]\right|^p\right)^{\frac{1}{p}}, \tag{4.20}$$

where, if p=2 we have the Euclidean distance. In general, p is selected depending on knowledge of data and the Euclidean distance value is adopted when this is unknown. Other possibilities for such measures exist depending on the magnitudes that have to be handled (Baseville 1989, McLachlan 1992).

In practice, there is a variety of ways in which the k nearest neighbors can be used to determine the class of $\boldsymbol{q}$. Although it can be straightforward to assign the class with the most elements among the k nearest neighbors, a more general notation is given in the form of distance weighted voting where neighbors are allowed to vote on the class of the query.

More precisely, let d^* denote the largest distance among the k nearest neighbors obtained by Eq. 4.18. Then, the class voting is calculated as,

$$V_c = \sum_{\substack{i \\ d \leq d^*}} \frac{1}{d\left(\boldsymbol{q}, \boldsymbol{x}_i\right)^n} 1\left(c = c_i\right), \tag{4.21}$$

for $i = 1, \ldots, k$.

When the element count is used instead of the class voting, the inverse is omitted and the vote is just the indicator function. However, it will often make sense to assign more weight to these neighbors depending on how close they are. In general, the lower the distance from $\boldsymbol{q}$ neighbor $\boldsymbol{x}_i$ is located the higher the importance of that neighbor. Finally, the class of $\boldsymbol{q}$ is calculated as

$$r_B\left(\boldsymbol{q}\right) = \arg\max_{c \in C}\left(V_c\right) \tag{4.22}$$

k-NN is equivalent to using class distributions of the form $\hat{f}_c(\boldsymbol{x}) = k_c / \left(N_c v_k\left(\boldsymbol{x}\right)\right)$, where $v_k\left(\boldsymbol{x}\right)$ is the volume of an ellipsoid centered at $\boldsymbol{x}$ and of radial d^*. This method and *Kernel density estimation* (Hand 1982) constitute the most common type of Bayes density estimates due to the intuitive observation that data does not have to be normally distributed in classes (as in LDA).

4.4 Learning Vector Quantization (LVQ)

Possibly, the *Learning Vector Quantization* (LVQ) (Kohonen et al. 1992) is considered the most intuitive method among all non-linear classifiers, as it is somehow similar to the nearest neighbor scheme. In the non-linear classifier, the principle is to build a blank structure (or a system composed of a network of elements) from scratch and configure it optimally during the learning phase so that input data can be classified with very reduced misclassification error.

Particularly in vector quantization classifiers, the elements of such a network are known as 'codebooks'. Each codebook is assigned to a class and is positioned at an arbitrary location in the space defined by the user. Then in the learning phase, the position of one codebook vector is changed at a time from one element of the input partition.

As usual, the elements come from the learning set $\mathcal{L}_t$. When the training stage is completed new elements, now coming from the testing set $\mathcal{L}_s$ are classified to determine the identification accuracy. In this case, however, the class is decided upon that of the *closest* codebook vectors, and not directly upon the elements of the learning set.

4.4.1 LVQ algorithms

Let us assume that a certain number of *codebook vectors*, denoted by $\boldsymbol{m}_i$, are placed into the same space of input data, where $\boldsymbol{m}_i \in \mathbb{R}^T$. In general, several codebook vectors are assigned to each class, and the input $\boldsymbol{x}$ is then decided to belong to the class $c_i \in [1, C)$ the nearest $\boldsymbol{m}_i$ belongs to. Let

$$r(\boldsymbol{q}) = \arg\min_{c_i} \|\boldsymbol{x} - \boldsymbol{m}_i\|, \tag{4.23}$$

define the nearest $\boldsymbol{m}_i$ to $\boldsymbol{x}$, denoted by $\boldsymbol{m}_c$. The values $\boldsymbol{m}_i$ that approximately minimize the misclassification error of $r(\boldsymbol{x})$ can be found asymptotically in one of the following learning processes.

4.4.1.1 The LVQ1

Let $\boldsymbol{x}(n)$ be an input sample from the learning set, and let $\boldsymbol{m}_c(n)$ represent sequences of $\boldsymbol{m}_c$ in the discrete-time domain. Starting with properly defined initial values, the following equations define the LVQ learning algorithm which has the effect of dividing up the input space in different class groups:

$$\begin{aligned}
\boldsymbol{m}_c(n+1) &= \boldsymbol{m}_c(n) + \alpha(n)[\boldsymbol{x}(n) - \boldsymbol{m}_c(n)] \\
&\quad \text{if } \boldsymbol{x} \text{ and } \boldsymbol{m}_c \text{ belongs to the same class,} \\
\boldsymbol{m}_c(n+1) &= \boldsymbol{m}_c(n) - \alpha(n)[\boldsymbol{x}(n) - \boldsymbol{m}_c(n)] \\
&\quad \text{if } \boldsymbol{x} \text{ and } \boldsymbol{m}_c \text{ belongs to the different classes,} \\
\boldsymbol{m}_i(n+1) &= \boldsymbol{m}_i(n) \text{ for } i \neq c
\end{aligned} \tag{4.24}$$

This algorithm corresponds to the most simple version of LVQ classifier. Here, $0 < \alpha(n) < 1$ is known as the learning parameter, and $\alpha(t)$ may be constant or decrease monotonically with n. In this version of LVQ, it is preferable to start with a value of α smaller than 0.1 (see subsection 4.4.1.5).

4.4.1.2 2nd Version (the LVQ2)

This method is identical to LVQ1 but in this case two codebook vectors $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$, the nearest to $\boldsymbol{x}$, are updated simultaneously in each step of the learning stage. One of them must belong to the correct class and the other to a wrong class, respectively. Moreover, $\boldsymbol{x}$ must fall into a window defined around the mid plane of $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$, respectively.

More precisely, let d_i and d_j be Euclidean distances of $\boldsymbol{x}$ from $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$, respectively; then $\boldsymbol{x}$ is said to fall into a window of relative width w if

$$\min\left(\frac{d_i}{d_j}, \frac{d_j}{d_i}\right) > s \tag{4.25}$$

where $s = (1-w)/(1+w)$. A relative width of -0.2 to 0.2 is recommendable in the following LVQ2 algorithm

$$\begin{aligned} \boldsymbol{m}_i(n+1) &= \boldsymbol{m}_i(n) - \alpha(n)\left[\boldsymbol{x}(n) - \boldsymbol{m}_i(n)\right], \\ \boldsymbol{m}_j(n+1) &= \boldsymbol{m}_j(n) + \alpha(n)\left[\boldsymbol{x}(n) + \boldsymbol{m}_j(n)\right], \end{aligned} \tag{4.26}$$

where $\boldsymbol{x}$ and $\boldsymbol{m}_j$ belong to the same class, $\boldsymbol{x}$ and $\boldsymbol{m}_i$ belong to different classes and the condition 4.25 must hold.

4.4.1.3 3rd Version (the LVQ3)

The LVQ2 algorithm is based on the idea of differentially shifting the decision borders towards the Bayes limits, while no attention is paid to what may happen to the location of the $\boldsymbol{m}_i$ during the execution process. Therefore, it seems necessary to include corrections that ensure the $\boldsymbol{m}_i$ continue approximating the class distributions.

Combining these ideas, LVQ2 is improved by also changing those codebook vectors between $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ that belong to the same class of $\boldsymbol{x}$. If we define these vectors as $\boldsymbol{m}_k$, the LVQ3 algorithm modifies their position in each step of the learning stage, additionally to $\boldsymbol{m}_i$, as

$$\boldsymbol{m}_k(n+1) = \boldsymbol{m}_k(n) + \epsilon\alpha(n)\left[\boldsymbol{x}(n) + \boldsymbol{m}_k(n)\right], \tag{4.27}$$

for all k such that

$$\min\left(\frac{d_i}{d_j}, \frac{d_j}{d_i}\right) < \min\left(\frac{d_i}{d_k}, \frac{d_k}{d_i}\right), \text{ and } \min\left(\frac{d_i}{d_j}, \frac{d_j}{d_i}\right) < \min\left(\frac{d_k}{d_j}, \frac{d_j}{d_k}\right). \tag{4.28}$$

Here, the idea is to use the class of an input element to modify the position of several codebook vectors of the same class. While the closest vectors are modified by means of 4.26,

the remaining ones are slightly changed by means of 4.27. Therefore, the parameter ϵ should be configured with a small value. Kohonen et al. (1995) proposed using values between 0.1 and 0.5. However, the optimal value of ϵ seems to depend on the size of the window w, being smaller for narrower windows. This method seems to be self-stabilizing (the optimal placement of the $\boldsymbol{m}_i$ does not change in continuous learning).

4.4.1.4 Differences between LVQ1, LVQ2 and LVQ3

The three presented options, namely, the LVQ1, the LVQ2 and the LVQ3, yield almost similar accuracies, although a different philosophy underlies each. The LVQ1 and LVQ3 are considered more robust, in the sense that the codebook vectors assume stationary values even after extended learning periods. For the LVQ1 the learning rate can be optimized for quick convergence. In the LVQ2, the relative distances of the codebook vectors from the class borders are optimized, although there is no guarantee for the codebook vectors being placed optimally to describe the forms of the class borders. Therefore, LVQ2 should only be used with a small value of the learning rate and a relatively low number of training steps.

4.4.1.5 Optimizing the Learning Rate

The optimized LVQ (OLVQ1) modifies the basic LVQ1 in such a way that each codebook vector $\boldsymbol{m}_i$ has its own individual learning rate $\alpha_i(n)$. Let c be the classification of sample $\boldsymbol{q}$ as defined in expression 4.23. Then, the following OLVQ1 learning algorithm has the effect of dividing up the space in different class groups, increasing the speed of convergence of the final configuration of codebook vectors

$$\begin{aligned}
\boldsymbol{m}_c(n+1) &= \boldsymbol{m}_c(n) + \alpha_c(n)\left[\boldsymbol{x}(n) - \boldsymbol{m}_c(n)\right] \\
&\quad \text{if } \boldsymbol{x} \text{ and } \boldsymbol{m}_c \text{ belongs to the same class,} \\
\boldsymbol{m}_c(n+1) &= \boldsymbol{m}_c(n) - \alpha_c(n)\left[\boldsymbol{x}(n) - \boldsymbol{m}_c(n)\right] \\
&\quad \text{if } \boldsymbol{x} \text{ and } \boldsymbol{m}_c \text{ belongs to the different classes,} \\
\boldsymbol{m}_i(n+1) &= \boldsymbol{m}_i(n) \text{ for } i \neq c
\end{aligned} \tag{4.29}$$

In order to determine the optimal $\alpha_i(n)$ for fastest convergence, Eq. 4.29 is rewritten in the form

$$\boldsymbol{m}_c(n+1) = \left[1 - s(n)\,\alpha_c(n)\right]\boldsymbol{m}_c(n) - s(n)\,\alpha_c(n)\,\boldsymbol{x}(n) \tag{4.30}$$

where $s(n) = +1$ if the classification of $\boldsymbol{x}$ is correct and $s(n) = -1$ if the classification is wrong. Notice, that $\boldsymbol{m}_c(n+1)$ contains traces from $\boldsymbol{x}(n)$ through the right term in Eq. 4.30, which

are scaled down by the factor $\alpha_c(n)$, and traces from the earlier sample $\boldsymbol{x}(n-1)$ through the equivalent right term of $\boldsymbol{m}_c(n)$ (if Eq. 4.30 is rewritten for n instead of $n+1$) scaled down by the factor $[1 - s(n)\,\alpha_c(n)]\,\alpha_c(n-1)$. Since it is known that the statistical accuracy of the learned codebook vector values is optimal when the effects of these corrections at the end of the learning period are of equal weight, these two scalings must be identical and, therefore,

$$\alpha_c(n) = [1 - s(n)\,\alpha_c(n)]\,\alpha_c(n-1)\,. \tag{4.31}$$

If this condition is made to hold for all n, it can be shown that the traces collected up to time n will be scaled down by an equal amount at the end and, thus the optimal values of $\alpha_i(n)$ are determined recursively by

$$\alpha_c(n) = \frac{\alpha_c(n-1)}{1 + s(n)\,\alpha_c(n-1)}. \tag{4.32}$$

With this provision, the learning stage is speeded up significantly, especially in the beginning, and the $\boldsymbol{m}_i$ find their approximate asymptotic values relatively quickly, even for rather high initial values of $\boldsymbol{\alpha}_i$, say, 0.3.

4.4.2 General Considerations

Although LVQ uses a nearest neighbor as the decision rule, it is a nonlinear classifier that tries to approximate class distributions by means of codebook vectors, which act as structure networks that must be configured to work properly. For this reason, the achievable accuracy and the required time for learning will always depend on the initialization of such a network. The latter is sometimes considered a rather complicated task if compared to the Bayes classifiers, as the learning rate or the stopping rule, among other parameters, are less evident in many cases.

4.4.2.1 Initialization of Codebook Vectors

In practical applications of pattern recognition, even when the class probabilities are quite different, experimenting with the same number of codebook vectors in each class is recommended first and, then, adjusting the quantity if necessary.

Since the class borders are represented by segments of mid planes between codebook vectors of neighboring classes, it may even seem a better strategy to set the average distance between adjacent codebook vectors equal on both sides of the borders. Then, if at least the class distributions were symmetric, the average shortest distance of codebook vectors would be the same in every class. However, because the final placement of vectors is not known

until the end of the process, the optimal number of vectors for every class and thus their distances cannot be determined before that. Therefore, the assignment of codebook vectors to the various classes can only be made iteratively.

Once the number of codebook vectors has been fixed, for their initial values one can use real values from samples in the training data set picked up from their respective classes. Since the codebook vectors should always remain inside their respective class domain, only the values corresponding to correctly classified elements can be accepted. A possible strategy may be to use some linear classifier (i.e the k-NN) for one tentative sample against all the other samples (by the leave-one-out method) in the training set and, then, accept this sample as codebook vector if this classification corresponds to the class of the sample.

It is also recommended to additionally compute the medians of the shortest distances between the initial codebook vectors of each class. Then, if the distances turn out to be very different for the different classes, new codebook vectors may be added to or old ones deleted from the deviating classes. In general, for a satisfactory approximation of the borders the medians of the shortest distances between the codebook vectors should be somewhat smaller than the standard deviations of the input samples in all the respective classes.

4.4.2.2 Learning and Stopping Rule

When using LVQ to build a classifier, it is recommended to start with the OLVQ1 in the learning stage for fast convergence. Normally, its accepted asymptotic recognition accuracy is achieved after completing a number of learning steps about 30 to 50 times the total number of codebook vectors. In general, the OLVQ1 alone may be enough for practical applications. However, in an attempt to improve recognition accuracy, one may continue the learning with either the basic LVQ1, the LVQ2 or the LVQ3, using a low initial value of the learning rate.

However, if the learning continues beyond an optimal network configuration, the algorithm loses the ability to generalize and, therefore, accuracy starts to decrease. It is, therefore convenient to stop after a certain number of steps. In the LVQ1, LVQ2 and LVQ3, the number of steps can range from 50 to 200 times the total number of codebook vectors. However, such a stopping rule can only be found by experience, and it also depends on the input data.

4.5 Conclusions

Some algorithms for the identification of classes and the validation of feature models have been reviewed in this chapter. In general, class distributions can be characterized by their

extension area or their decision boundaries in the feature space.

Linear methods exploit the first strategy. In this case, estimated distributions depend on the position of inputs from the learning set, and decision rules rely on the proximity of the elements to be classified to these inputs so, in general, the configuration will not be complex and the classification algorithms have low computational cost. However, they lack the capacity to fit to very complex distributions of data. In this case, it is much better to use the second approach.

Although it may seem that the second group of classifiers achieve high classification rates, a major problem is the lack of intuition to configure their learning stage so that the method converges to an optimal solution. For example, in methods such as Neural-Networks, CART or LVQ there is no single "rule-of-thumb" that specifies the necessary number of layers and perceptrons, the depth of the tree or the number of codebook vectors that are necessary for the application (Breiman et al. 1984, Kohonen et al. 1995). In the context of this work, the use of a representative coordinate feature system are aimed at reducing the complexity of the classification task.

Part III

Contributions

Chapter 5

On LDB-based Pattern Recognition

5.1 Introduction

Habitually, the process of looking structure patterns in data requires the extraction and validation of the different class features. Although the preprocessing tasks of fish identification and aging differ significantly due to the nature of their data, the feature extraction and classification algorithms reviewed in previous chapters may be chosen for both the analysis of contour irregularities and the intensity profile.

The Best-Basis paradigm and the LDB algorithm provide an interesting framework for automating the selection of features in classification tasks, as long as their limitations regarding the reduction of dimensionality are solved. In particular, if we manage to find a reduced vector presenting a clearly separated space distribution, not only will the classification accuracy will improve considerably but also the computational cost will benefit from the use of very simple classification methods.

Therefore, this chapter is dedicated to presenting our new LDB design, which addresses three issues: the evaluation of coefficients, the discriminant measures and the search algorithm.

The first problem focuses on the evaluation of separability. In this sense, we believe that evaluating probability distributions is more important than considering the amount of signal energy. Figure 5.1 illustrates this problem, by means of two example distributions, which assume two class Normal densities of equal variance but different class mean. A coefficient may be selected on the basis of presenting high energy when its class distributions really overlap each other, while another coefficient may be discarded because of its low energy, although it possess great separation capabilities.

Secondly, a mechanism for controlling the correlation between discriminant measurements

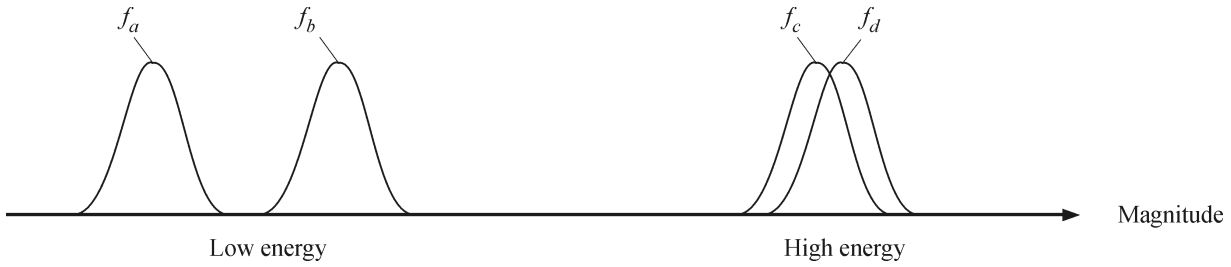


Figure 5.1: Limitations of the use of the energy magnitude in the evaluation of discrimination power. f_a and f_b are the density distributions of two well separated classes, whereas f_c and f_d overlap each other, despite having much more energy

and classification efficiency must exist. An entire range, characterized by its lower and upper bounds, must also be able to represent the boundaries of classification accuracy, if some degree of generalization is to be attained. Commonly used measures (Baseville 1989, McLachlan 1992) do not control this behavior, so a framework is necessary to provide normalization and reduce this subjectivity as far as possible.

The last issue focuses on the divide-and-conquer algorithm used in the generation of admissible trees. The discrimination power that represents a node is obtained considering all its elements (see section 3.3.3.1 in chapter 3), but the contribution of single elements to this measure may be more representative compared to other members in this node. In this situation, a criterion based on the consideration of single coefficients may offer more robust behavior.

To address all these problems we propose the flow chain illustrated in Fig. 5.2. Our new LDB, which comprises all the stages from DWPT to divide-and-conquer, will be part of both fish identification and aging applications. However, because of the different nature of their data the preprocessing path varies. On one hand, we have developed a normalization framework for closed contours that reduces the time-translation effects of the wavelet transform. The proposed prefiltering scheme for intensity profiles, on the other hand, tries to address the problems of image contrast and growth modulation, discussed in chapter 2. As such, the stages associated to this path constitute a feature-based approach of fish age estimation.

Contour normalization and growth demodulation are described in sections 5.2 and 5.5, respectively. A brief introduction of non-parametric density estimation is included in section 5.3, to justify the generalization capabilities of the proposed framework. Our proposal is called *'Density Local Discriminant Bases* (DLDB) and is reviewed in greater detail in section 5.4. Final remarks regarding our contributions are discussed in section 5.6.

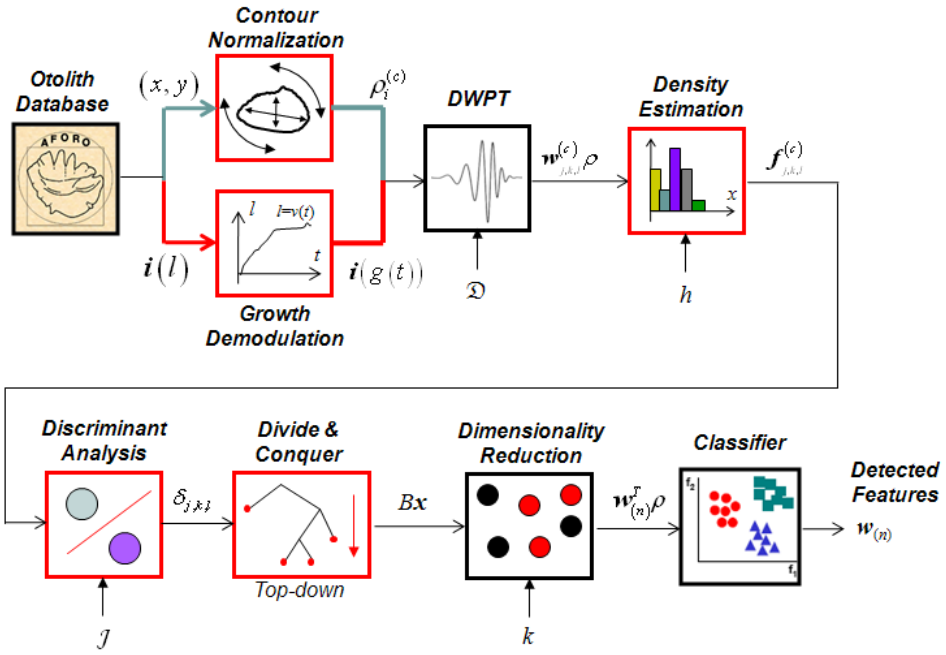


Figure 5.2: Detailed diagram of the proposed feature extraction scheme in this thesis. Our proposals involve methods for both fish identification and aging applications (blocks highlighted in red) and extend the general diagram depicted in Fig. 3.13. The new LDB implements the DWPT, density estimation, discriminant analysis and a divide-and-conquer algorithm. Dimensionality reduction is developed manually.

5.2 1st. Proposition: Size, Rotation and Translation Normalization

Whether expressed in polar or cartesian notation, contour data generally comes in four basic forms of geometric distortion: translation, rotation, scaling and shearing. Since the impact of this distortion can be severe in otolith-based fish identification applications, using descriptors that resit this geometric attack in the contour shape is of crucial relevance to compare otoliths of several species. In this sense, some authors have used either morphological criteria or numerical transformation to avoid these problems but, in practice, each otolith must be normalized with respect to its specific intrinsic shape properties (Lombarte et al. 2003, Cadrin and Friedland 1999, Piera et al. 2005).

One way of standardizing the otolith contour is provided by the ellipse corresponding to the first component of the discrete Fourier series (Kuhl and Giardina 1981). As illustrated in Fig. 5.3, the goal is to modify contour rotation, first sample position and radial magnitude according to the main ellipse of the otolith.

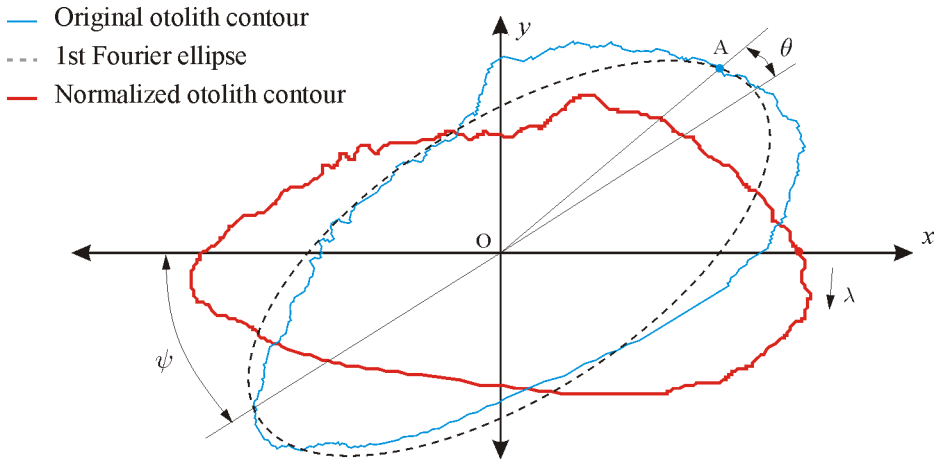


Figure 5.3: Normalization by the semi-major axis of the first Fourier harmonic ellipse. Distance (OA) denotes the radial of this semi-major axis. Normalization is performed rotating the otolith contour by ψ and shifting all coordinate point by the amount of λ

More precisely, let a_1, b_1, c_1 and d_1, be the first coefficients corresponding to a non-normalized otolith contour with radial $\rho[n]$ and angle $\theta[n]$. Assuming N absolute coordinate points, that is $\{x,[n],y[n]\}$ for $n \in [0, N-1)$, we have,

$$\begin{aligned} a_1 &= \tfrac{2}{N}\sum_{n=0}^{N-1}\cos\left(2\pi\tfrac{n}{N}\right)(x[n]-\mu_x), \qquad b_1 = \tfrac{2}{N}\sum_{n=0}^{N-1}\sin\left(2\pi\tfrac{n}{N}\right)(x[n]-\mu_x), \\ c_1 &= \tfrac{2}{N}\sum_{n=0}^{N-1}\cos\left(2\pi\tfrac{n}{N}\right)(y[n]-\mu_y), \qquad d_1 = \tfrac{2}{N}\sum_{n=0}^{N-1}\sin\left(2\pi\tfrac{n}{N}\right)y([n]-\mu_y), \end{aligned} \tag{5.1}$$

where μ_x and μ_y denote contour center point, defined as the mean of x- and y-axis coordinate points, respectively. The normalized coefficients, denoted by '*', can be determined as,

$$\begin{bmatrix} a_1^* & c_1^* \\ b_1^* & d_1^* \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{bmatrix} \begin{bmatrix} a_1 & c_1 \\ b_1 & d_1 \end{bmatrix}, \tag{5.2}$$

where θ_1 is the angle between the semi-major axis of the first harmonic ellipse regarding the reference axis of the original contour, which can be calculated as

$$\theta_1 = \frac{1}{2}\arctan\left[\frac{2(a_1b_1 + c_1d_1)}{a_1^2 + c_1^2 - b_1^2 - d_1^2}\right]. \tag{5.3}$$

Thus, the rotation angle necessary for alignment with this axis, ψ, is readily obtained as,

$$\psi = \arctan\frac{c_1^*}{a_1^*}. \tag{5.4}$$

Full normalization is obtained by scaling the radials relatively to the otolith size and shifting by λ, were $\lambda = \theta_1 N/2\pi$ denotes position of the first sample with respect to the original reference expressed on a number-of-sample basis. These operations are in addition to those of angle rotation. Therefore, full contour is characterized by the normalized contour samples

$$\rho_{\mathcal{N}}[n] = \frac{\rho[n-\lambda]}{\rho_0} \tag{5.5}$$

and

$$\theta_{\mathcal{N}}[n] = \theta[n-\lambda] - \psi, \tag{5.6}$$

expressed on a polar coordinate system, and $\rho_0 = (a_1^{*2} + b_1^{*2})^{1/2}$ corresponds to the radial of the semi-major axis.

5.3 Density Estimation for Local Feature Description

Traditionally, two different approaches for estimating distributions in data exist: *nonparametric* and *parametric*. The focus of each other differs remarkably. In the latter, given a family function, $f(\cdot|\theta)$ such as the two-parameter Normal family $N(\mu, \sigma^2)$ where $\theta = (\mu, \sigma^2)$, the emphasis is on obtaining the best estimator, $\hat{\theta}$, of θ which makes $\boldsymbol{f}$. In general, parametric approaches do not require large amounts of data but they do involve a lot of prior work gathering previous expert knowledge and information in order to produce reliable templates.

The emphasis of nonparametric estimation, on the contrary, is on directly estimating a version $\hat{\boldsymbol{f}}$ of the true density function $\boldsymbol{f}$ from data, which thus eliminates the need for a model specification. Despite the large amount of data necessary, this approach is generally very intuitive, flexible and often provides estimations close to true densities (Scott 1992).

5.3.1 Histograms

The most simple way to develop nonparametric estimation within LDB consists in building *density histograms* for each one of the different basis functions in the wavelet tree. The histogram is completely determined by the sample $\{x_1, \ldots, x_{N_c}\}$ and a choice of mesh $\{t_m, -\infty < m < \infty\}$.

Let $G_m = [t_m, t_{m+1})$ denote the m-th bin of fixed width $h = t_{m+1} - t_m$ for all m. A density histogram uses building blocks of height $1/(nh)$ so that each block has an area equal

to $1/n$. Let v_m denote the bin count of the m-th cell, that is, the number of samples falling into G_m. Then, the estimated density functions at each point in the tree is defined as

$$\hat{f}\left[x\right] = \frac{v_m}{hN_c} = \frac{1}{hN_c}\sum_{i=1}^{N_c} I_{[G_m)}\left(x_i\right), \tag{5.7}$$

where $I_A\left(x\right) = 1$ if $x \in A$, and 0 otherwise.

Clearly, the usefulness of the estimation of $\boldsymbol{f}$ depends on an optimal choice of the bin width h, very often referred to as the "smoothing" parameter since it controls the amount of smoothness of the profile in $\boldsymbol{f}$. The desire to compare different estimators and to attempt to identify the best one assumes the specification of a criterion that can be optimized. Although very simple ideas, such as the Sturges' number-of-bin rule (Sturges 1926), apply and are configured in many software packages by default, data-based optimization methods try to minimize the *integrated square error* (ISE) and/or its average value (MISE) which are calculated estimating the *bias* and *variance* components. When the L_2-norm is used, this value is defined as

$$\text{MISE} = \int E\left[\hat{\boldsymbol{f}} - \boldsymbol{f}\right]^2 dx. \tag{5.8}$$

However, knowing the true density function $\boldsymbol{f}$ is considered to be one of the most difficult tasks a priori. In this sense, approximations are made assuming the existence of derivatives of $\boldsymbol{f}$ and employing a variety of techniques known as cross-validation algorithms. Following this norm, the asymptotic optimal width is defined as a function of the first derivative roughness of $\boldsymbol{f}$ as

$$h^* = \left(\frac{6}{\phi\left(\boldsymbol{f}'\right)}\right)^{1/3} N^{-1/3}, \tag{5.9}$$

where $\boldsymbol{\phi}$ denotes the roughness estimator adopted for the L_2- norm of ϕ

$$\phi\left(\boldsymbol{f}'\right) = \int f'\left(x\right)^2 dx. \tag{5.10}$$

and N is the total number of samples used in the construction of the histogram (Scott 1979, Freedman and Diaconis 1981). With this estimator of the roughness and bin width, the optimal asymptotic MISE (AMISE) is known to be

$$\text{AMISE}^* = \left(3/4\right)^{2/3} \phi\left(\boldsymbol{f}'\right)^{1/3} N^{-2/3} \tag{5.11}$$

This result has led to the width of some standard distributions, such as the Normal case with $\boldsymbol{f} = N(\mu, \sigma^2)$, being fixed. It can be shown that $\phi(\boldsymbol{f}') = 1/(4\sqrt{\pi}\sigma^3)^{1/3}$. Hence, from Eq. 5.9 we obtain

$$h^* = \left(24\sqrt{\pi}\sigma^3/N\right)^{1/3}. \tag{5.12}$$

This rule constitutes a reference for constructing histograms. One can first use the sample standard deviation $\hat{\sigma}$ in (5.12) and estimate the Normal bin width as

$$\hat{h}_N \approx 3.5\hat{\sigma}N^{-1/3}, \tag{5.13}$$

and then adapt it to more realistic distributions, such as lognormal or t-student.

Another possibility suggested by Terrel and Scott (1985) is to use $\hat{h}_{OS} = 3.729\sigma N^{-1/3}$, which constitutes the mesh for over-smoothed histograms of fixed bin width. Any of these rules can be modified multiplying them by the skewness factor

$$\beta_1(\sigma) = \frac{2^{1/3}\sigma}{e^{5\sigma^2/4}\left(\sigma^2+2\right)^{1/3}\left(e^{\sigma^2}-1\right)^{1/2}}, \tag{5.14}$$

when data is skewed or the kurtosis factor

$$\beta_2 = 3 + \frac{(d-2)^{1/2}\, B\left(\frac{3}{2}, \frac{2d+3}{2}\right)}{d^{3/2} B\left(\frac{1}{2}, \frac{d+1}{2}\right)^2}, \tag{5.15}$$

if data is heavy-tailed. Here, d denotes degrees of freedom of the t distribution, and $B(a,b) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$, with Γ being the *gamma* function.

5.3.1.1 Bin Width Selection by Cross-validation

More robust methods, however, calibrate the bin width automatically. They are known as cross-validation algorithms. The goal is to produce bin widths that minimize the MISE by reusing data (Scott and Terrell 1987). In the calculation of MISE a version of the roughness for finite samples is estimated.

Let $\hat{f}[x]$ be the constructed histogram with the data at hand using bin width h. An estimation of $f'[x]$ is then available based on a finite difference of the histogram at the midpoints of bins G_m and G_{m+1}, namely, $\hat{f}'[x] = [v_{m+1} - v_m]/(nh^2)$, so a potential estimate of $\phi(\boldsymbol{f}')$ is

$$\hat{\phi}_1\left(\hat{\boldsymbol{f}}'\right) = h\sum_m \left(\hat{f}'[x]\right)^2 = \frac{1}{N^2h^3}\sum_m [v_{m+1} - v_m]^2. \tag{5.16}$$

However, this estimation of roughness is biased. Indeed, it can be demonstrated that

$$\mathrm{E}\left[\hat{\phi}_1\right] = \phi\left(\hat{\boldsymbol{f}}'\right) + 2/\left(nh^3\right) + O\left(h\right). \tag{5.17}$$

With an optimal smoothing, the term $2/\left(Nh^3\right)$ converges to $\phi\left(\hat{\boldsymbol{f}}'\right)/3$ leading to a biased estimator which is too large (by a factor of a third) so that $\frac{3}{4}\hat{\phi}_1$ is an asymptotically unbiased estimation of $\phi\left(\hat{\boldsymbol{f}}'\right)$. Subtracting the quantity $2/\left(Nh^3\right)$ to Eq. 5.16 and introducing the result into the AMISE expression 5.11 gives a *biased cross-validation* (BCV) estimate of the MISE:

$$\mathrm{BCV}\left(h\right) = \frac{5}{6Nh} + \frac{1}{12N^2h}\sum_m \left(v_{m+1} - v_m\right)^2, \tag{5.18}$$

where v_m and v_{m+1} is recomputed for different meshes of h (the bin origin t_0 remains fixed, for definiteness). The BCV bin width, $\hat{h}_{BCV}$, is defined to be the minimizer

$$\hat{h}_{BCV} = \arg\min_h \mathrm{BCV}\left(h\right), \tag{5.19}$$

subject to the constraint $h \leq h_{OS}$. The bias simply refers to the fact that the error estimation used by this algorithm is biased to the true MISE.

An unbiased version can be obtained if the ISE, which is defined as the integrated square error

$$\mathrm{ISE} = \int \left[\hat{\boldsymbol{f}} - \boldsymbol{f}\right]^2 dx = \phi\left(\hat{\boldsymbol{f}}\right) - 2\int \hat{f}\left(x\right) f\left(x\right) dx + \phi\left(\boldsymbol{f}\right), \tag{5.20}$$

is used within the minimization process of the cross-validation algorithm, instead of the MISE. In this case, it can be demonstrated that the minimizer does not depend on the unknown quantity $\phi\left(\boldsymbol{f}\right)$ and that the second integral can be rewritten as $\mathrm{E}\left[\hat{f}\left(x_i\right)\right]$, where the expectation is with respect to the point of evaluation, x_i, and not over the whole sample $x_1, \ldots, x_N$ (see Rudemo 1982 for the details). This suggests removing one data point to use the remaining N-1 points to construct histograms that can serve as estimators of $\mathrm{E}\left[\hat{f}\left(x_i\right)\right]$.

Let $\hat{f}_{-i}\left[x\right]$ denote the histogram, which is constructed excluding data point x_i. The quality of fit of this point is evaluated as $\hat{f}_{-i}\left[x\right] = \left(v_m - 1\right)/\left[\left(n-1\right)h\right]$ if $x_i \in G_m$. Then, repeating this process N times (once for each data point) and averaging the result yields $\mathrm{E}\left[\hat{f}\left(x_i\right)\right]$ which can be substituted in Eq. 5.20 to derive the *unbiased cross-validation* function

$$\begin{aligned} \mathrm{UCV}\left(h\right) &= \phi\left(\hat{\boldsymbol{f}}\right) - \frac{2}{N}\sum_{i=1}^{N} \hat{f}_{-i}\left(x_i\right) \\ &= \frac{2}{\left(N-1\right)h} - \frac{N+1}{N^2\left(N-1\right)h}\sum_m v_m^2. \end{aligned} \tag{5.21}$$

In general, the BCV and UCV are the most common references to estimate the efficiency of the histogram (Scott 1992). The UCV is much noisier in complex distributions, whereas the minimizer of the BCV is generally biased toward larger bin widths. So the UCV is more correct on average, despite the fact the presence of strong local minimizers constitutes one of the main problems to set an optimal value for h.

The lesson that can be derived from practical experience is that oversmothing, BCV and UCV should be examined simultaneously, even with very large data sets. Specifically, the UCV and BCV curves should be plotted on comparable scales, marking the location of the upper bound provided by the oversmoothed width. However, since the variability of input data is, somehow, represented within the validation method of identification accuracy, when using some of the methods described in section 4.2.1, it is reasonable to reconsider the bin width at this step.

5.4 2nd. Proposition: Density Local Discriminant Basis

We are now able to describe the different modifications of the LDB algorithm (described in chapter 3) in greater detail. In this section, the wavelet coefficients constitute the main input data points.

Our proposition is based on two fundamental probabilistic properties that, we believe, discriminant measures should accomplish in order to reflect a more realistic classification capacity of the coefficients and, therefore, make the generalization through different application data possible. In our opinion, this requirement is met by relating discrimination and classification accuracy to one another. The construction of density curves should be restricted within boundaries and the development of measures should not loose coherence in relation to identification rates. Our principal hypothesis are the following:

Hypothesis 1. Normalized Density. *Let $\boldsymbol{f}^{(c)}$ denote the distribution at some position (j,k,l) of the wavelet tree in class c and let $\delta \in \mathbb{R}$ be its discrimination power. Assume that a classifier, r, by which a classification accuracy $1-\hat{R}\left(r\left(x\right)\right)$ is obtained, exists. Then, generalization restrict density estimation to normalized functions:*

$$\sum_{m} f^{(c)}\left[x\right] = 1. \tag{5.22}$$

This is the most basic property of density functions in probabilistic theory and means that this condition should hold for every single position (j,k,l) in the wavelet tree and class c. This is an essential requirement in order to control the values of discriminant measures

within specified boundaries (the next subsection looks at this issue in greater detail).

Hypothesis 2. Discrimination-Accuracy Correlation. *Let* $\mathcal{A}_c \equiv \{x : a_c \leq x \leq b_c\}$ *represent the partition* (or space) *of samples of the* c *class in a particular position of the wavelet tree, which has discrimination power whose distribution* δ*, and whose class distributions are* $\boldsymbol{f}_c$*. Then, the condition*

$$\begin{cases} \delta = L, & \text{if } \mathcal{A}_i \cap \mathcal{A}_j = \mathcal{U} \\ L < \delta < U, & \text{if } \mathcal{U} \supset (\mathcal{A}_i \cap \mathcal{A}_j) \supset \emptyset \\ \delta = U, & \text{if } \mathcal{A}_i \bigcap \mathcal{A}_j = \emptyset \end{cases} \tag{5.23}$$

must hold for $i \neq j$*, where* L *and* U *correspond to lower and upper bounds of measure* δ*, respectively,* and $\mathcal{U}$ *denotes the universal set.*

This rule forces two different class partitions to produce discrimination values predefined within boundaries, L and U. In addition, this value should grow inversely to the intersecting area of their joint operation. Assume that we use L=0 and U=1. Then, the proposed behavior of δ is consistent with the obvious fact that completely disjoint class spaces allow the correct classification of all elements ($\hat{R}(r) = 0$ if $\delta = 1$) and, on the contrary, useless features ($\delta = 0$) assign the classes at random with a misclassification accuracy of $\hat{R}(r) = 1 - \frac{1}{C}$.

On the other hand, how δ should vary within L and U when $\mathcal{A}_i \bigcap \mathcal{A}_j \neq \emptyset$ and $\mathcal{A}_i \bigcap \mathcal{A}_j \neq \mathcal{U}$ is a question that deserves careful attention. From our point of view, this variation should be as linear as possible with the identification rate $\hat{R}(r)$, which should grow in $0 > \hat{R}(r) > 1 - \frac{1}{C}$, for $1 > \delta > 0$.

In practice, however, there are many factors involved in this behavior related to the particular limitations in which the $\boldsymbol{f}^{(c)}$ are estimated, the discriminant measure function, or even the method used to validate classification accuracy. More specifically, such factors tackle the number of input samples available, the size of feature vectors or the classification method, among others. As such, it will always be necessary to contrast the second hypothesis by comparing discrimination and accuracy.

To make LDB work towards this principle, we considered three aspects, two involving the generalization of discriminant measures and another involving the manner in which coefficients are found and selected to obtain an admissible tree. The calculation of accurate density distributions has already been described in the previous section with several methods.

In what follows, both the construction of normalized discriminant functions and the search topic is considered. Although the search of features is not directly related to the concept of data generalization, it plays an important role in the physical interpretation of selected

features. The proposed methodology will be referred to as the *Density Local Discriminant Bases* (DLDB).

5.4.1 Towards the Standardization of δ

A possible way to construct bounded discriminant measures with normalized densities and achieve input generalization, is to complement the Bhattacharyya affinity measure as

$$\jmath\left(a,b\right)=1-\sum_{x}\left(\hat{f}^{(a)}\left[x\right]\hat{f}^{(b)}\left[x\right]\right)^{1/2} \tag{5.24}$$

where a and b denote two different classes.

Clearly, this measure has a minimum value, $\jmath=0$, for exactly distributed classes ($\hat{\boldsymbol{f}}^{(a)}=\hat{\boldsymbol{f}}^{(b)}$, for all m and all $a\neq b$) and a maximum value, $\jmath=1$, for completely disjoint distributions (each m has at least one $\hat{\boldsymbol{f}}^{(c)}=0$). Additionally, since $\sum_m \hat{f}^{(c)}\left[x\right]=1$ and $\hat{f}\left[x\right]<1$ for all m, the value computed in Eq. 5.24 is ensured to grow inversely to the overlapping range of both classes, so a straightforward modification of the LDB algorithm is to replace the energy maps (Eq. 3.84) by the density estimate in Eq. 5.7, and use Eq. 5.24 to determine the discrimination value at each position in the DWPT tree.

For notation convenience, let us change the WPT coefficient $\boldsymbol{w}_{j,k,l}\boldsymbol{x}_i^{(c)}$, corresponding to the i-th signal of class c at position (j,k,l), into $x_i^{(c)}\left(j,k,l\right)$ and assume that $f_{j,k,l}^{(c)}\left[x\right]$ is the histogram representing its class density distribution $f_{j,k,l}\left(x\right)$. To extend the measure to several classes one possibility is to consider the weighted addition of all $\binom{C}{2}$ pair combinations

$$\delta_{j,k,l}=\binom{C}{2}^{-1}\sum_{a=1}^{C-1}\sum_{b=a+1}^{C}\jmath_{j,k,l}\left(a,b\right), \tag{5.25}$$

where $\jmath_{j,k,l}\left(a,b\right)$ is calculated at position (j,k,l). Despite these changes, the resulting LDB algorithm (the LDB2) remains almost similar to that of the original version (see subsection 3.3.3.1):

- **Step 0:** *Choose a dictionary of orthonormal bases* $\mathfrak{D}$ (*i.e. specify* QMF's *for a wavelet packet dictionary or decide to use either the local cosine dictionary or the local sine dictionary*). *Specify the maximum level of decomposition* J *and a type of 'bounded' discriminant measure* $\jmath$.

- **Step 1:** *Construct the normalized density histograms for each basis function,* $\hat{\boldsymbol{f}}_{j,k,l}^{(c)}$, *and class* $c=1$, ..., C *and determine discriminant measures* $\delta_{j,k,l}$.

- **Step 2:** *Set* $B_{J,k} = A_{J,k}$ *and* $\Delta_{J,k} = \delta_{J,k}$ *for* $k = 0, \ldots, 2^J - 1$ (where $\delta_{j,k} = \sum_{l=0}^{2^{n_0-j}-1} \delta_{j,k,l}$).

- **Step 3:** *Determine the best subspace* $B_{j,k}$ *for* $j = J-1, \ldots, 0$*;* $k = 0, \ldots, 2^j - 1$ *by the following rule:*

  - **Set** $\Delta_{j,k} = \delta_{j,k}$.
  - **If** $\Delta_{j,k} \geq \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$,
  - **then** $B_{j,k} = A_{j,k}$,
  - **else** $B_{j,k} = B_{j+1,2k} \cup B_{j+1,2k+1}$ and set $\Delta_{j,k} = \Delta_{j+1,2k} + \Delta_{j+1,2k+1}$.

- **Step 4:** *Order the selected basis functions* $\boldsymbol{w}_n$ (*stored in* $B_{0,0}$) *by their power of discrimination.*

- **Step 5:** *Use* k ($\leq N$) *most discriminant basis functions for constructing classifiers.*

As nodes are evaluated by accumulating the discriminant values of all the coefficients, some problems arise. A node showing only a few particularly high discriminant coefficients may be discarded because it contains too many useless coefficients, while another node may be selected simply because it shows a certain average, though not necessarily display outstanding discrimination coefficients.

Another problem deals with the search strategy. The comparison of nodes starts at the highest level (or scale) and ends up evaluating, $W_{0,0}$, against the tree configuration stored at the previous step. For this reason, the divide-and-conquer algorithm is said to use a *bottom-up* strategy. If the final decision, $B_{0,0} = W_{0,0}$ is taken the whole search would be in vain since no relevant information is obtained with the selection of the space containing the original signals.

5.4.2 Node Selection Combined with Top-down Search

In the following divide-and-conquer algorithm a *top-down* strategy for the LDB is described. In this case, the node coefficients with particularly high discrimination power have major priority in the calculation of the admissible tree, which means that the node is selected according to its coefficient with highest $\delta_{j,k,l}$, for all $j = 0, \ldots, J$ and $k = 0, \ldots, 2^j$, denoted by l.

Let $\mathcal{B} \equiv \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N\}$ represent the LDB, we are looking for, and let $\mathcal{T}$ be initialized to the DWPT tree containing all of its nodes,

$$\mathcal{T} = \bigcup_j \left(\bigcup_k W_{j,k} \right)$$

for $j = 0, \ldots, J$ and $k = 0, \ldots, 2^j - 1$. Let $\mathcal{C}$ be defined as a set variable containing the part of the DWPT tree unfolded from node (u,v) and downwards,

$$\mathcal{C} = \bigcup_j \left(\bigcup_k W_{j,k} \right)$$

where $j = u, \ldots, J$ and $k = 2^{j-u}v, \ldots, 2^{j-u}(v+1) - 1$. Here, $W_{u,v}$ is a working variable of the search algorithm. Then, given a set of training signals of known class $\left\{ x_i^{(c)} \right\}_{i=1}^{N_c}$, the following steps define the new LDB algorithm

Algorithm 5.1: Density Local Discriminant Basis (DLDB)

- **Step 0:** *Choose a dictionary of orthonormal bases* $\mathfrak{D}$ (*i.e. specify* QMF's *for a wavelet packet dictionary or decide to use either the local cosine dictionary or the local sine dictionary*). *Initialize* $\mathcal{B}$ *as an empty set of nodes* ($\mathcal{B} = \emptyset$) *where the selected spaces will be stored. Specify the maximum level of decomposition* J *and a 'bounded' discriminant measure* $\jmath$.

- **Step 1:** *Construct the normalized density histograms for each base function,* $\hat{\boldsymbol{f}}_{j,k,l}^{(c)}$, *and class* $c = 1, ..., C$; *and estimate their associated discrimination value* $\delta_{j,k,l}$.

- **Step 2:** *Repeat the following process until the tree is completely pruned* ($\mathcal{T} = \emptyset$):

  - **Set** $W_{u,v}$ as,

    $$W_{u,v} = \arg \max_{W_{j,k} \in \mathcal{T}} (\delta_{j,k,l})$$

    and update $\mathcal{C}$ accordingly.
  - **If** $\mathcal{B} \cap \mathcal{C} = \emptyset$,
  - **then** $\mathcal{B} = \mathcal{B} \cup W_{u,v}$ and $\mathcal{T} = \mathcal{T} - \mathcal{C}$,
  - **else** $\mathcal{T} = \mathcal{T} - W_{u,v}$.

- **Step 3:** *Order the selected basis functions* $\boldsymbol{w}_n$ (*stored in* $\mathcal{B}$) *by their power of discrimination* δ_n.

- **Step 4:** *Use* k ($\leq N$) *most discriminant basis functions for constructing classifiers.*

5.5 3th. Proposition: Automatic Growth Demodulation (AGD)

It was already mentioned in chapter 2 that otolith growth is influenced by factors such as habitat, water temperature, fish diet or light, among others. In this case, the main problem when estimating fish age consists in how precise the Von Bertanlanffy parameters (Eq. 2.27) can be estimated, while trying to avoid the negative effects of image contrast variation. In general, it is unlikely that all specimens obey the same growth parameters, just as the parametric method of section 2.4.2 suggests.

In our opinion, other ways of demodulating age structures which are much more consistent with reality rely on the use of individualized growth functions, which assume the existence of an independent growth variability for each otolith. That is, a different β_i for each specimen using the same growth model template $\hat{\boldsymbol{v}}(t,\beta)$. The philosophy underlying this approach is more consistent with the fact that events during the fish's life, combined with genetic factors, have diverse influence on growth behavior, suggesting that the growth curve should be adapted for each single fish.

Recall that the fish growth was defined as $l = v(t[n])$ for the intensity profile vector $i(l) = i[n]$, for $n \in \mathbb{Z}$, where l and t represents otolith radial length and fish age, respectively. Our proposed AGD algorithm attempts to estimate the growth of the fish automatically. The goal is to find the optimal growth parameters $\beta^* = \{L_\infty, t_0, k\}$ that obtain the most regular periodicity of age structures in the demodulated profile $\boldsymbol{i}(\hat{v}(t,\beta))$, but without knowing the position of year marks l_i .

The autocorrelation function is often used in signal processing as a means of detecting repeating patterns which have been buried under noise (Proakis and Manolakis 1996). In this sense, the short-time auto correlation function (STACF), which is equivalent to the autocorrelation of a windowed sequence (Rabiner and Schafer 2007), is used to represent the periodicity at several positions along the demodulated intensity profile. That is

$$\tau_{m,\beta}(k) = \frac{1}{\sigma^2}\sum_{n=0}^{N-k}\left(i_{m,\beta}[n]-\mu\right)\left(i_{m,\beta}[n+k]-\mu\right) \quad \text{for } k < N, \tag{5.26}$$

where $i_{m,\beta}[n] = \boldsymbol{i}(\hat{v}(t,\beta))\, w[n-m]$ denotes the piece of demodulated signal using growth parameters β, with mean μ and variance σ^2, for a rectangular window $w[n] = \mathbf{1}_{[0,W)}$.

Note that all non-zero elements in $\boldsymbol{\tau}$ are relative to the window, positioned at m, and we have a total of M different representations. So the first largest lag, that actually corresponds to an oscillating year period, should present similar relative distance inside the windows if a constant periodicity is to be present along the profile (otherwise, the lags would be diffused,

indicating thus that the demodulation parameters β have not been selected properly). A natural way to estimate how the lags are concentrated in the windows is to construct density functions with all them, and then use an entropy measure.

Let $\hat{f}_\beta(u)$ denote an estimation for this density implemented by a histogram

$$\hat{f}_\beta(u) = \frac{1}{hM} \sum_m I_{[k_j, k_{j+1})}(u_{m,\beta}). \tag{5.27}$$

Here, $k_j = hj$ establish the bin origins for a total number of J bins, with $k_j < W$ and width h, whereas $u_{m,\beta}$ corresponds to the relative position of the largest lag to window m using demodulation β, determined as

$$u_{m,\beta} = \arg\max_{k<N} [\tau_{m,\beta}(k)] - m \quad \text{for } \tau(\cdot) > \gamma, \tag{5.28}$$

where γ denotes a signal threshold. Finally, the entropy measure

$$H(\beta) = -\sum_j f_\beta(u) \log_2 f_\beta(u), \tag{5.29}$$

allows to estimate the concentration of lags: the lowest value indicates that all bins are zero, except for one, meaning thus that all lags are positioned at the same place, or equivalently, that the year periodicity is regular along the demodulated profile. At this point, we are in situation to provide a formal definition for what an optimal demodulation is.

Definition 5.1. Optimal Fish Growth Estimation: *Given a modulated sequence* $i(l) = i[n]$, *of length* N ($n \in \mathbb{Z}$), *where* $l = v(t[n], \beta)$ *corresponds to the modulation law, the* β^* *that minimizes the entropy measure* $H(\cdot)$,

$$\beta^* = \arg\min_\beta [H(\beta)], \tag{5.30}$$

provides the most possible regular periodicity of oscillatory structures in the demodulated signal $i(\boldsymbol{v})$.

Such kind of problems are normally solved by means of classical *min-search* algorithms (Spendley et al. 1962, Nelder and Mead 1965). Although the Nelder-Mead algorithm does not require many step evaluations to converge to a solution, in general it is difficult that β^* reaches the global minimum of (5.30). Furthermore, sometimes it may have undesired effects, such as converging to non-stationary points or following search directions orthogonal to the gradient of the objective function. So, if we do not restrict the space of solutions it may become computationally infeasible (Han and Neumann 2006). Therefore, a two-parameter version of growth model functions $v(t) = L_\infty(1 - \exp[-kt])$ is used in our experiments.

5.6 Conclusions

Three new image and signal processing methods have been proposed in the context of fish otolith applications and have been formalized and fully described in this chapter. These include 1) a normalization method for closed contours, 2) a method for demodulating the periodicity of oscillatory structures in age signals (the AGD), and 3) a new LDB feature extraction algorithm: the DLDB.

The goal of these algorithms is to provide efficient coordinates of representation for signal classification. The normalization of contours is not only important to generalize among otoliths of different size and rotation but also to avoid the translation effect of the wavelet packet transform, which is embedded within the DLDB method. Also, obtaining a regular periodicity of age structures via signal demodulation is important to improve the estimation of age in fishes. In this sense, the operation principle of AGD in the automatic demodulation of growth structures has been presented.

The DLDB modifies considerably the way coefficients are evaluated compared to the original LDB method: a completely different and normalized discriminant measure and search criterion in the feature space has been used. This approach constitutes the main core of both applications, otolith-based fish identification and age estimation, and offers a new framework to automating the selection of features in automatic algorithms of pattern recognition.

In principle, density-based techniques and bounded discriminant measures should provide greater consistency in the evaluation of descriptor models, regardless the application we are using. This property is necessary to obtain good generalization behavior. The last part of this document (chapters 6 and 7) will be devoted to showing our results towards this property and to proposing guidelines for future development.

Part IV

Results and Conclusions

Chapter 6

Fish Identification and Age Estimation Results

6.1 Introduction

In order to test the proposed feature extraction tools, results for otolith-based fish identification and aging applications are presented in this chapter. The main purpose of our experimentation will be to evaluate the different preprocessing frameworks presented previously and demonstrate the generalization capabilities of our feature extraction scheme, the DLDB.

All the algorithms have been developed in MATLAB code and C language. The only exception is the LVQ method where a version from the LVQ_PAK software, obtained from the the author's web page (Kohonen et al. 1995), was used. The work of this thesis has been developed in the context of the AFORO3D project (Análisis de FORmas de Otolitos 3D, Centro Superior de Investigacions Científicas, CSIC) and the AFISA project from the European Union (Automated FISH Aging, nº. 044132) in collaboration with several research institutions: the French Research Institute for Exploitation of the Sea (IFREMER), the Institute of Marine Research (IMR, Tromso-Norway), the Marine Research Institute of Iceland (MRI), the Centre for the Environment, Fisheries and Aquaculture Science (CEFAS, England), among others.

Some results included here have been published and presented in conferences (Soria et al. 2008, 2009, Soria and Parisi-Baradad 2011) and part of these tools, specially, the DLDB and the demodulation tools, were included in the TNPC v4.0 NOESIS software package (www.noesis.com).

In fish identification, we will evaluate the performance of different descriptive an analytical

tools: the Discrete Fourier Transform (DFT), the standard Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), which will be compared with the original Local Discriminant Bases (LDB) and our proposal, the Density Local Discriminant Bases (DLDB). In all the experiments, the otolith contour points form our Standard Euclidean Basis (STD).

For this purpose, a two-class classification experiment has been set up with two fish species of different genera, the *Merluccius merluccius* and *Gadus morhua* (common hake and cod, respectively), whose otoliths are very different from each other. From a biological point of view this experiment may be of very little interest since identification of both species could be developed easily using the external fish shape. But from the perspective of method implementation it will bring more insight into the use of LDB methods. In particular, their capacity in interpreting the physical structures of otolith irregularities and determining their location on the image, the way to configure them or the improvements introduced with DLDB, will be demonstrated. Two classification methods will be used in this experiment: the LVQ approach and the k-Nearest Neighbor method.

Another two-class identification experiment will study the inter-specific and intra-specific relations of *Merluccius* spp. from different geographical locations in the world. Generally speaking, *Merluccius* spp. are distributed in the Northern and Southern Hemisphere, on both sides of the Atlantic Ocean, throughout the eastern Pacific from the USA-Canadian border to Cape Horn, along the Mediterranean and Black seas, and off New Zealand (Fig 6.1).

In this case, eight inter-specific and intra-specific test pair groups were chosen. The species used in the intra-specific tests were: 1) *M. merluccius* (Linnareus 1758); 2) *M. bilinearis* (Mitchill 1818); 3) *M. gayi* (Guichenot 1848); and 4) *M. productus* (Ayres 1855). The test pairs for the inter-specific tests were: 1) *M. albidus* (Mitchill 1814) and *M. bilinearis* from north-west Atlantic; 2) *M. capensis* (Castelnau 1861) and *M. paradoxus* (Franca 1960) from south-east Atlantic. 3) *M. polli* and *M. senegalensis* from central-east Atlantic (Cadenat 1950); and 4) *M. gayi* and *M. australis* from south-east Pacific (Hutton 1872). These pair groups were chosen because they have different population characteristics, and are of very important economic value in general. The hakes from the intra-specific pairs are found in two different world areas and, therefore, allow us to find environmental factors attributed to the species, whereas in the inter-specific groups the two hake species show overlapped geographical distributions (sympatric species).

The experiment will also serve to observe the behavior of DLDB under a more difficult identification problem, and in particular the sympatric species, whose species inhabit the same geographical region and are quite similar each other. In this case, the otolith is a necessary element in order to succeed in their identification.

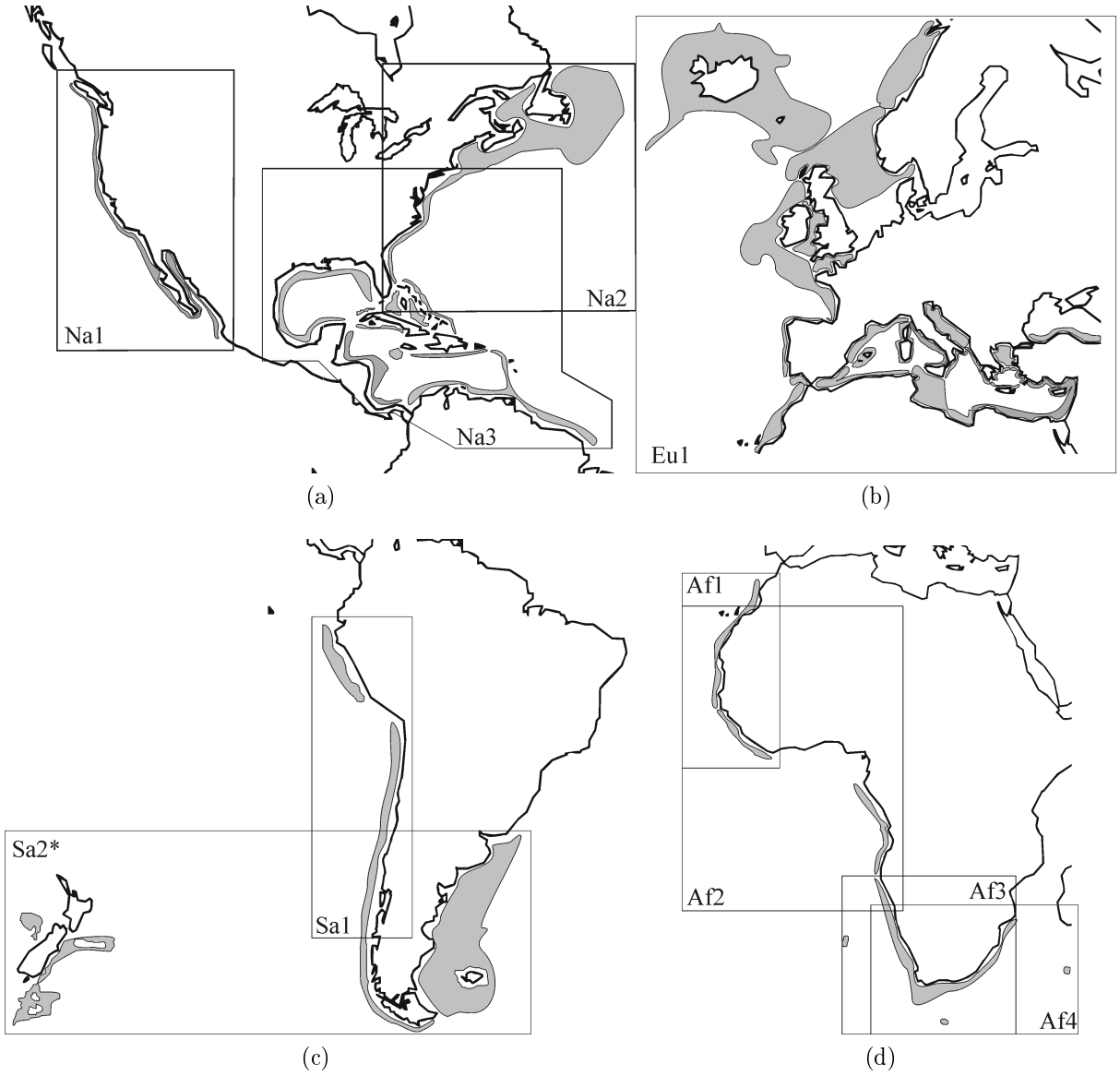
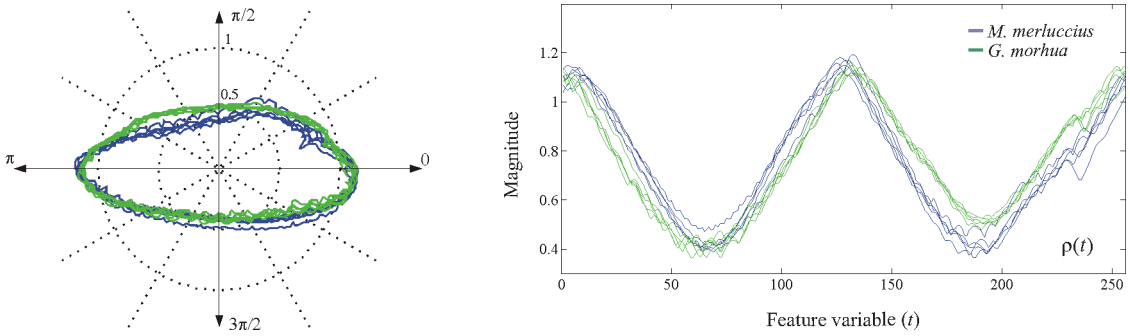


Figure 6.1: World distribution of the ten *Merluccius* species considered in this work (based on Lloris et al. 2005). a) North American Pacific and Atlantic (Na1: *M. productus*, Na2: *M. bilinearis*). b) Europe (Eu1: *M. merluccius*). c) South American Pacific and Atlantic (Sa1: *M. gayi*, Sa2: *M. australis*). d) Africa (Af1: *M. senegalensis*, Af2: *M.polli*, Af3: *M. capensis*, Af4: *M. paradoxus*).

In aging applications, the goal is to compare manual and automated preprocessing methods of trending extraction and growth demodulation, respectively, and study their impact in the estimation of fish age. For this purpose, we have set up an experiment consisting in the identification of five year classes of plaice otoliths: up to two year-old fishes (1st. class);



(a) (b)

Figure 6.2: Normalized otolith contour examples from *Merluccius merluccius* (blue) and *Gadus morhua* (green). a) Polar representation b) Clockwise representation of the radials. The coordinates have been resampled to 256 points in order to use the 1D version of the DLDB

three, four and five year-old fishes (classes 2-4); and six years or over (5th. class).

According to the previous chapters there are a total of four methods involved in the preprocessing of the intensity profile, but the DLDB and the LVQ remain intact as the main core of our general pattern recognition system (see Fig. 5.2). For the manual methods we use the MCE method for contrast cancellation and the SGD for demodulating fish growth, whereas the AGF and AGD respectively, are equivalent algorithms for automatic methodologies (the reader is referred to chapters 2 and 5 for the details).

In this sense, of relevant importance is the gain that is obtained in the estimation of age by each one of the preprocessing tools. Therefore, we first consider feature extraction alone with no kind of preprocessing method. Then we extend the evaluation to contrast cancellation and growth demodulation independently, in their manual and automatic fashion. And finally, we evaluate the effect of implementing the two approaches together.

6.2 Application by Fish Identification

6.2.1 Analysis of *M. merluccius* and *G. morhua*

A total of 129 otoliths from the CEFAS collection (http://cefas.dedra.gov.uk) were used in our study. The radial profile presents significant differences at the dorso-caudal margin where the shape is less irregular in *Merluccius merluccius* and the radials are larger than *Gadus morhua* (Fig. 6.2). Here, the x-y points have been extracted and normalized using the discrete elliptic Fourier approach (section 5.2) resampling to exactly 256 points, since dyadic lengths allow fast algorithms to be implemented for both DFT and DWPT (Beylkin 1993),

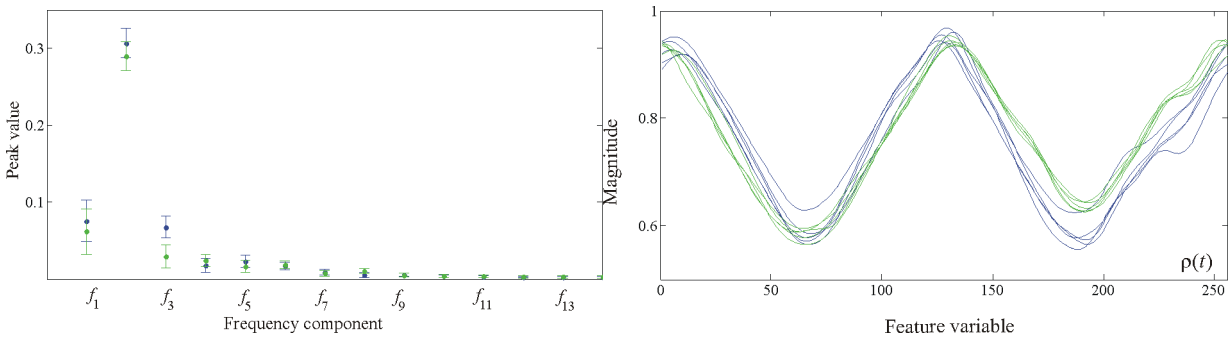


Figure 6.3: Fourier analysis of *Merluccius merluccius* (blue) and *Gadus morhua* (green) specimens. a) the fourteen lowest frequency components (offset component f_0, not included). b) Radial reconstruction using $f_0 - f_8$

respectively.

6.2.1.1 Results by Fourier Description

The offset component of the radial f_0 was considerable: about 0.75 ± 0.01 for *M. merluccius* and 0.77 ± 0.02 for *G. morhua*. The highest magnitude (a peak value of approximately 0.33) was obtained for two oscillating periods f_2, whereas the remaining features were below 0.1(Fig. 6.3a). Features f_3 and f_8 presented fully separated class distributions, each with respect to both species. This observation is relevant because it reveals that both class species present highly separated features. The signal reconstruction using the first eight features ($f_0 - f_7$, see Fig. 6.3b) demonstrates that most of the signal energy is concentrated at lower frequencies, with small oscillating transients located at the right half of the signals.

6.2.1.2 Results by PCA and LDA

A PCA analysis comparing *M. merluccius* and *G. morhua* otoliths was also carried out. Results are represented in scatter maps for three combinations of the best ten components (higher eigenvalues): 1) $f_{(1)}$ and f_2; 2) $f_{(1)}$ and $f_{(3)}$; and 3) $f_{(1)}$ through $f_{(10)}$, reduced to a two-dimensional representation thanks to the Sammon map approach (Fig. 6.4).

The behavior of STD and DFT was compared in each case. In general, STD showed better performance than DFT in all cases, since the clusters of both classes appear slightly more separated, making this type of representation more suitable than DFT for detecting all the specimens of both classes, perhaps by simply using a Bayes classifier.

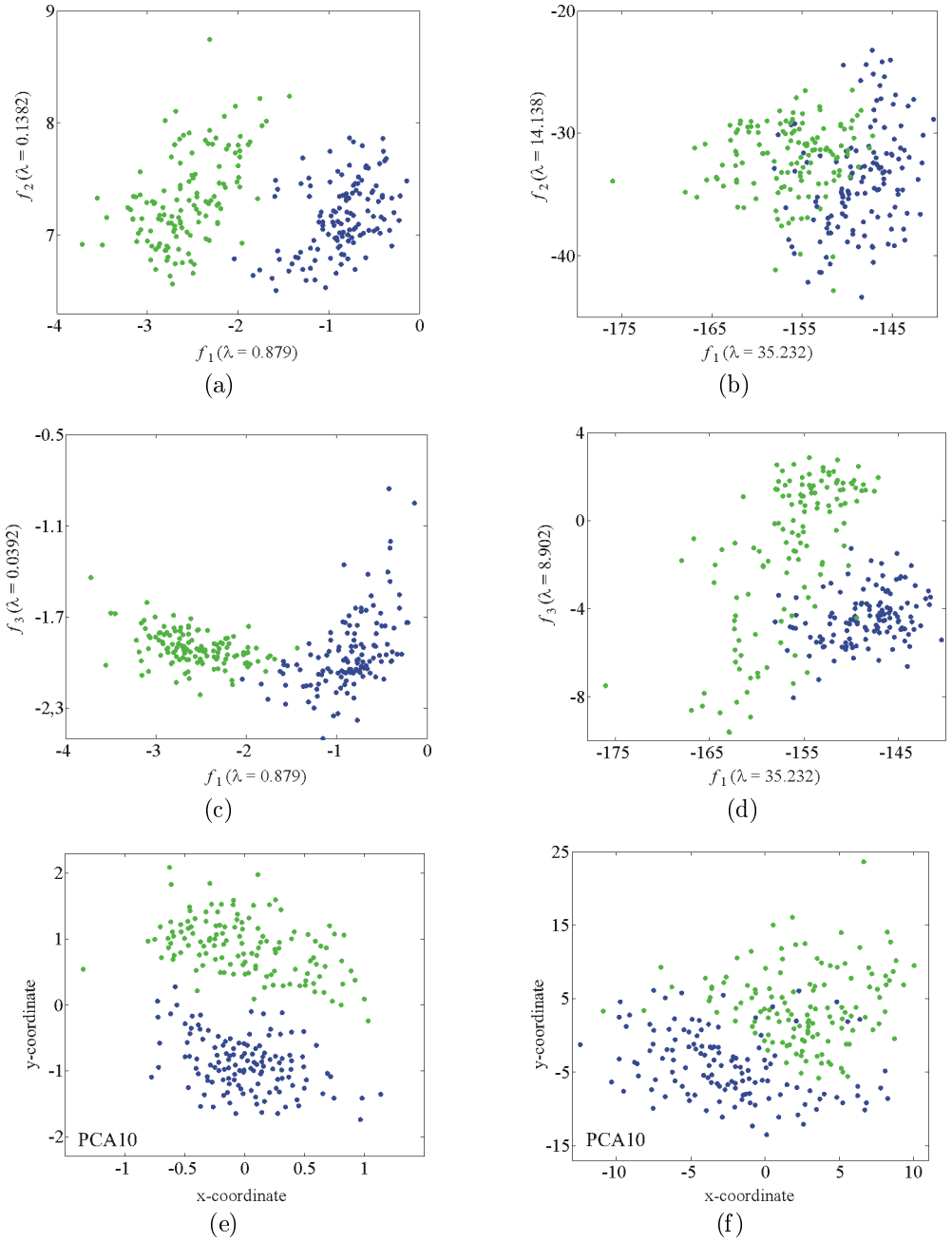


Figure 6.4: PCA results for *M. merluccius* (blue) and *G. morhua* (green). The distribution of STD and DFT coefficients is represented by means of scatter maps (left and right column, respectively). Results are presented for features a-b) $f_1 - f_2$; c-d) $f_1 - f_3$ and e-f) the best 10 PCA coefficients (represented in $\mathbb{R}^2$ by means of the Sammon map dimension reduction technique)

Input domain	λ	Coefficients			
		M. merluccius		*G. morhua*	
		Min.	Max.	Min.	Max.
STD	18578	-85.34	-85.30	-83.70	-83.66
DFT	17.25	0.399	0.466	0.497	0.579

Table 6.1: Results by LDA analysis

In STD, using either the first (Fig 6.4a) or third configuration (6.4d) of features is recommended, whereas in DFT, class distributions are a little more overlapped in all cases. The results also demonstrate that PCA eigenvalues are not representative measures of class separation. As an example, the eigenvalue of the best DFT component $f_{(1)}$ is $\lambda = 35.232$, while the best eigenvalue of the STD component is $\lambda = 0.879$ and offers better cluster separation.

A similar analysis was developed by means of LDA. In this case, the classes restrict the number of output features to only C-1 coefficients for each otolith, a single coefficient in this case. Since the value range of this coefficient does not overlap (Table 6.1), all the *M. merluccius* and *G. morhua* should be correctly identified in the classification phase either by STD or DFT. The eigenvalue is also uncontrolled and can not be used for evaluation purposes (STD and DFT have different eigenvalues, $\lambda = 18578$ and $\lambda = 17.25$ respectively, when both have the same discrimination). Therefore, the LDA eigenvectors are useful in the characterization of both species.

6.2.1.3 LDB and DLDB analysis

M. merluccius and *G. morhua* were also analyzed by means of LDB and DLDB. All information about results is provided through figures 6.5 - 6.8 and tables 6.2-6.3.

A total of $NJ = 256 \times 8 = 2048$ coefficients were calculated for each otolith contour using the *reverse biorthogonal 3.1* wavelets (rbior3.1). This dictionary was chosen because the five most relevant coefficients obtained the highest discrimination results with DLDB, $\delta(B\boldsymbol{x}) = \sum_{n=1}^{5} \delta_{(n)} = 4.989$ out of a maximum 5 (Table 6.2) .

LDB was configured with the symmetric Kullback Liebler criterion, whereas DLDB used the Battacharyya criterion to evaluate density histograms of $m = 9$ bins, which was chosen according to Sturge's rule. This configuration allowed discriminant measures to be normalized within the range [0,1).

The divide-and-conquer algorithm of both LDB and DLDB generated their respective admissible trees (Fig. 6.5). While the black cells correspond to the selected nodes, the previously mentioned five important coefficients are highlighted by means of red marks.

DLDB		LDB		DLDB		LDB	
Dictionary	$\delta(B\boldsymbol{x})$	Dictionary	$\delta(B\boldsymbol{x})$	Dictionary	$\delta(B\boldsymbol{x})$	Dictionary	$\delta(B\boldsymbol{x})$
Daubechies				*Biorthogonal*			
db31	**4.9287**	db36	**0.0580**	bior3.9	**4.9296**	bior3.1	**1.4933**
db35	4.7989	db42	0.0328	bior3.5	4.9296	bior3.5	0.7446
db40	4.3242	db4	0.0326	bior2.8	4.3189	bior6.8	0.1099
db8	3.2380	db5	0.0325	bior6.8	3.2448	bior1.1	0.0321
db7	3.1107	db33	0.0321	bior4.4	3.2035	bior4.4	0.0045
Symlets				*Reverse Biorthogonal*			
sym7	**4.9331**	sym27	**0.0328**	rbior3.1	**4.989**	rbior3.1	**0.2688**
sym11	4.9266	sym11	0.0327	rbior1.5	4.7833	rbior4.4	0.1315
sym2	3.9150	sym18	0.0326	rbior3.9	4.1210	rbior2.8	0.1044
sym30	3.4440	sym9	0.0325	rbior4.4	3.3694	rbior3.3	0.0170
sym14	3.2477	sym10	0.0324	rbior5.5	3.2572	rbior1.3	0.0045
Coiflets							
coif4	**3.5000**	coif5	**0.0326**	coif2	3.4474	coif4	0.0325
coif5	3.4744	coif3	0.0325	coif3	3.4445	coif1	0.0325
coif1	3.4588	coif2	0.0325				

Table 6.2: Evaluation of different dictionaries with DLDB and LDB. These include the *Daubechies*, *Symlets*, *Coiflets*, *Biorthogonal* and *Reverse Biorthogonal*. Measure values are provided considering the five best features in each dictionary. The best results are in bold format

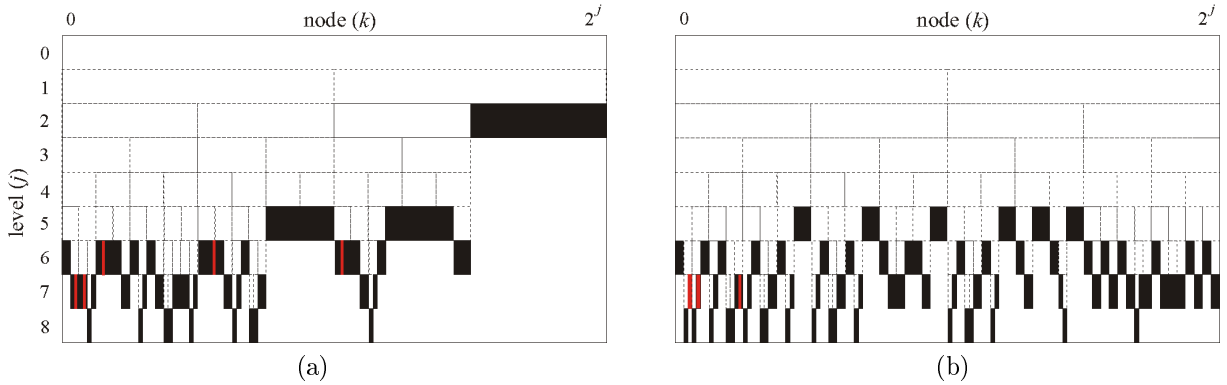


Figure 6.5: Admissible trees for the *M. merluccius* - *G. morhua* experiment using the *reverse biorthogonal 3.1* wavelets. Best nodes are shown in black and the best five coefficients are shown in red for a) the DLDB, and b) the LDB.

LDB				
Feature	$\delta_{(t)}$	Atom	$\boldsymbol{u_{min} - u - u_{max}}$	$\boldsymbol{\xi_{min} - \xi - \xi_{max}}$
$f_{(1)}$	0.1216	$\boldsymbol{\psi}_{7,3,1}$	0.332 -1.708 - 2.907	0.163 - 0.42 - 0.678
$f_{(2)}$	0.0561	$\boldsymbol{\psi}_{7,3,0}$	3.447 - 4.854 - 5.70	0.163 - 0.42 - 0.678
$f_{(3)}$	0.0374	$\boldsymbol{\psi}_{7,15,0}$	3.447 - 4.995 - 6.082	0.186 - 0.434 - 0.683
$f_{(4)}$	0.0332	$\boldsymbol{\psi}_{7,5,1}$	0.241 - 1.623 - 2.96	0.171 - 0.427 - 0.684
$f_{(5)}$	0.0205	$\boldsymbol{\psi}_{7,5,0}$	3.377 - 4.757 - 6.051	0.171 - 0.427 - 0.684
DLDB				
$f_{(1)}$	1.0	$\boldsymbol{\psi}_{7,5,1}$	0.241 - 1.623 - 2.96	0.171 - 0.427 - 0.684
$f_{(2)}$	1.0	$\boldsymbol{\psi}_{7,3,1}$	0.332 -1.708 - 2.907	0.163 - 0.42 - 0.678
$f_{(3)}$	1.0	$\boldsymbol{\psi}_{6,32,3}$	6.235 - 0.419 - 1.661	0.73 - 0.854 - 0.978
$f_{(4)}$	1.0	$\boldsymbol{\psi}_{6,16,3}$	6.251 - 0.594 - 1.616	0.412 - 0.51 - 0.607
$f_{(5)}$	0.989	$\boldsymbol{\psi}_{6,4,3}$	0.096 - 0.594 - 1.316	0.196 - 0.437 - 0.678

Table 6.3: Information on the location and frequency properties of selected features. The centering position of the atom (and its variability) is in radians. Discrete frequency values have been normalized using $2\xi/N$

Both LDB and DLDB provided the tree index $\gamma = (j, k, l)$ and discrimination power δ_γ of their best 256 basis functions. Bear in mind that the triplet $\gamma = (j, k, l)$ can also be seen as a way for indexing the structure patterns associated to the basis function $\psi_\gamma[n]$ in the full tree, which can be obtained after setting the corresponding cell $c_{j,k}[l]$ to one (with the remaining coefficients set to zero), reconstructing with the inverse DWPT (IDWT), and using Eq. 3.11 and 3.12.

This allowed the main feature parameters associated to the five coefficients to be determined (Table 6.3) , and their associated basis functions to be represented (Fig. 6.6). Since the irregularities of the rbior3.1 wavelets contain several artifacts, the variances of both the center position and frequency $\{u, \xi\}$ were generally large, thus generating waveforms of large time support and complex spectrograms of multiple frequency bands.

This information was used to highlight the otolith contour parts which present the most remarkable differences between both species (Fig. 6.7). According to Table 6.3, the best LDB features are located on the dorso-rostral area, the dorso-caudal margin and the ventral area, but nothing can be concluded about their relevance, despite the discriminant measurement of $\delta = 0.1216$.

As for DLDB, the Battacharyya criterion and the use of normalized density functions permitted to include "hot bars" and use colors for indicating the importance of features over the contours. The best features were located at the dorso-rostral margin and dorso-caudal area, but in this case the maximum possible discrimination power ($\delta = 1$) was obtained. This

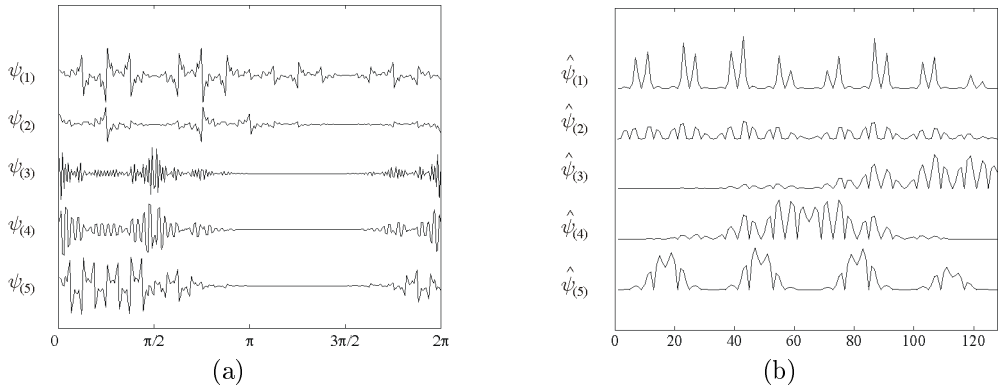


Figure 6.6: Waveform patterns for the five basis functions selected by the DLDB using the rbior3.1 dictionary. a) These representations correspond to the atoms in table 6.3. b) Their frequency spectrum is calculated by means of DFT.

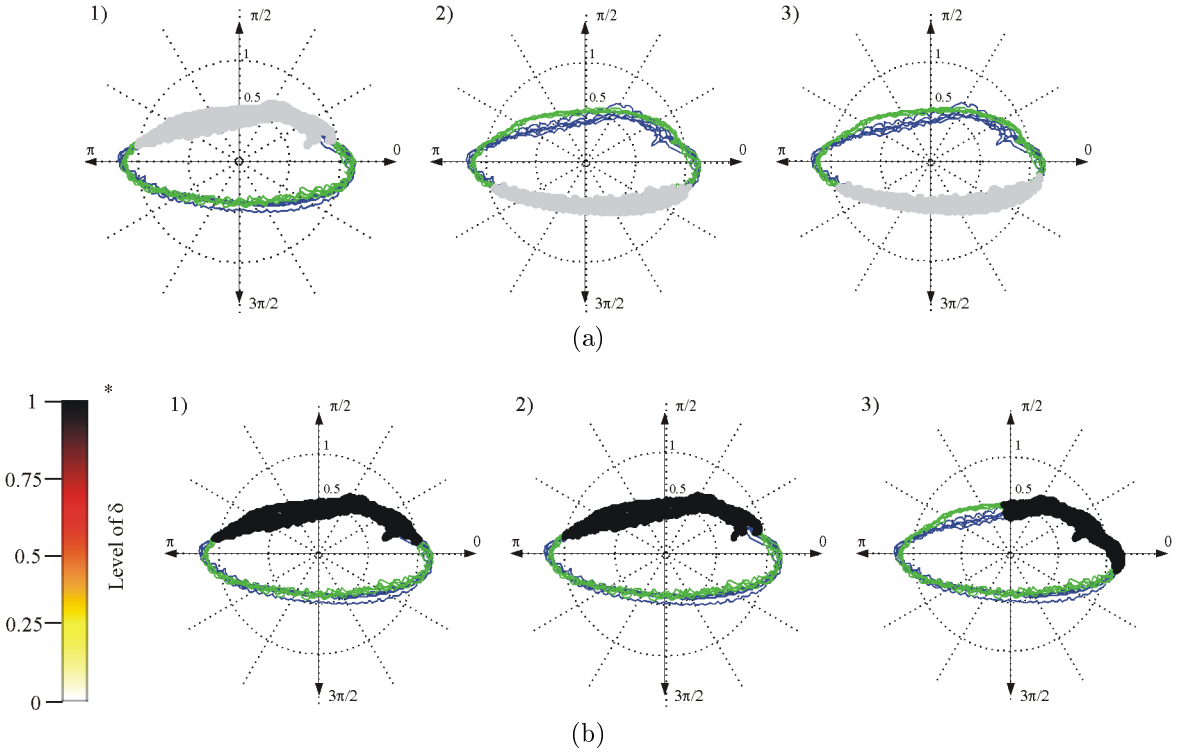


Figure 6.7: Detection of differences between cod (green) and hake (blue) at the otolith contour. a) LDB, b) DLDB. Maximum discrimination is obtained for the rbior3.1 wavelets at the dorso-caudal and dorso-rostral margins.

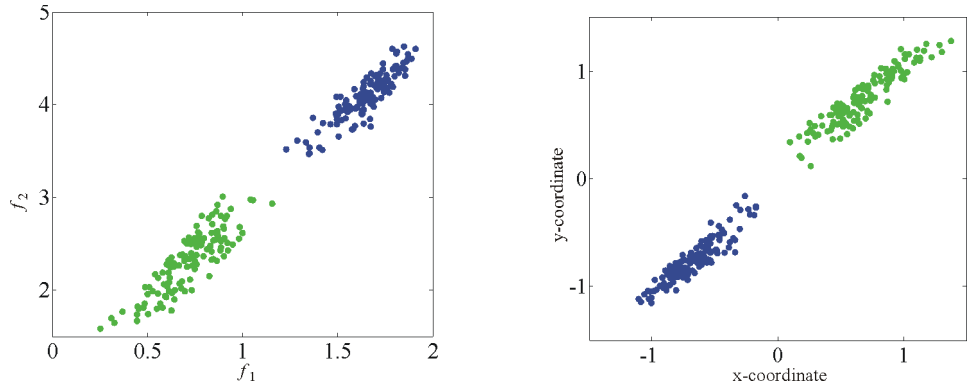


Figure 6.8: Scatter maps for cod (green) and hake (blue) features using DLDB. Representations are provided for a) the best two coefficients and, b) the best five coefficients (represented in $\mathbb{R}^2$ by means of the Sammon map dimension reduction technique)

fact indicates that both species are fully separable, as confirmed by the scatter plots in Fig. 6.8. The position of clusters suggests that DLDB is by far, more robust than PCA.

Therefore, it can be concluded that the otolith radial characterizes the main differences of both fish groups. Results from LDB and DLDB suggest that the morphology in *G. morhua* is more oblong and has relatively larger radials compared to the lancelolated shape in *Merluccius* otoliths at the dorso-rostral and dorso-caudal margin, as concluded from Fig. 6.2.

Contrary to what is expected, Fourier descriptors are not more efficient than STD, despite their eigenvalue measures (see subsection 6.2.1.2). The fact that this behavior occurs with otoliths of two easy distinguishable fish species is striking, and may question the general applicability of Fourier descriptors and PCA for pattern recognition purposes.

6.2.1.4 Classification of *M. merluccius* and *G. morhua*

Finally, the previously selected feature models and the proposed DLDB were validated by estimating their classification accuracy and correlating this value with discriminant measurements. For this purpose, the k-NN linear classifier and the LVQ nonlinear classifier were used.

In the learning stage, the Euclidean Distance was set as the default metric and basic rule to infer the classes for both classifiers. As for k-NN, the number of neighbors was set to $k = 10$ and the element-count approach was used to assign the classes to each sample in the testing set. The number of samples in both species was considered sufficient and the *hold-out* method was used to validate estimated rates. The proportion of samples used in the testing set was 0.3 and the number of repetitions was set to 50.

Feature	$f_{(1)}$	$f_{(1-2)}$	$f_{(1-5)}$	$f_{(1-10)}$
LVQ (Non-linear)				
DLDB on STD	99.56 (0.99)	100 (0)	100 (0)	100 (0)
LDB on STD	100 (0)	99.68 (0.58)	99.79 (0.48)	99.8 (6.2)
DFT on STD	71.98 (9.14)	77.95 (6.99)	96.97 (2.91)	96.75 (3.10)
PCA on STD	97.78 (2.04)	98.9 (1.63)	99.84 (0.81)	99.97 (0.19)
PCA on DFT	81.12 (6.51)	86.46 (5.16)	96.97 (2.64)	96.62 (2.81)
LDA on STD	100 (0)	-	-	-
LDA on DFT	100 (0)	-	-	-
k-NN				
DLDB on STD	99.81 (0.47)	100 (0)	100 (0)	98.84 (1.59)
LDB on STD	100 (0)	89.78 (5.1)	93.61 (3.62)	94.05 (3.81)
DFT on STD	75.31 (8.15)	56.86 (8.26)	63.39 (8.49)	57.72 (9.40)
PCA on STD	97.72 (2.45)	57.08 (10.18)	56.76 (8.10)	46.21 (8.97)
PCA on DFT	79.75 (6.36)	65.03 (7.26)	50.77 (8.68)	49.89 (9.37)
LDA on STD	100 (0)	-	-	-
LDA on DFT	100 (0)	-	-	-

Table 6.4: Classification results of *M. merluccius* and *G. morhua*. Identification accuracy is specified in %. The standard deviation calculated by means of the 'hold-out' method is also specified (in brackets).

Table 6.4 shows the classification accuracy for different feature extraction schemes and classification methods (from now on $\hat{R}$ is referred to as the correct classifications, for notation convenience). The best results were obtained for LDA, LDB and DLDB, all identifying almost all specimens. DLDB clearly outperforms standard PCA implemented on DFT, PCA on STD) and DFT on STD, obtaining almost full accuracy with only one feature in both classifiers. Other standard methods, such as DFT on STD or PCA on DFT, needed up to ten coefficients to reach a similar level with the LVQ classifier.

In general, LVQ showed more robust performance than k-NN in this experiment. In the latter, the accuracy decreases as more coefficients are included in the feature vector, except for DLDB and LDB, where accuracy values were almost similar. A possible reason of this robustness is the searching criteria included in both LDBs which try to select the additional features next in degree of importance.

Clearly, the use of DFT and PCA yielded the poorest results with the k-NN classifier. While the accuracy of PCA on STD was 97.72%, the results of PCA on DFT as well as DFT on STD were inferior to 80% when a single coefficient was used. Although the classification efficiency of LDA has been similar to both LDBs, it is evident that the possibility to detect and interpret the results in terms of the structure patterns of the otolith contour makes LDB

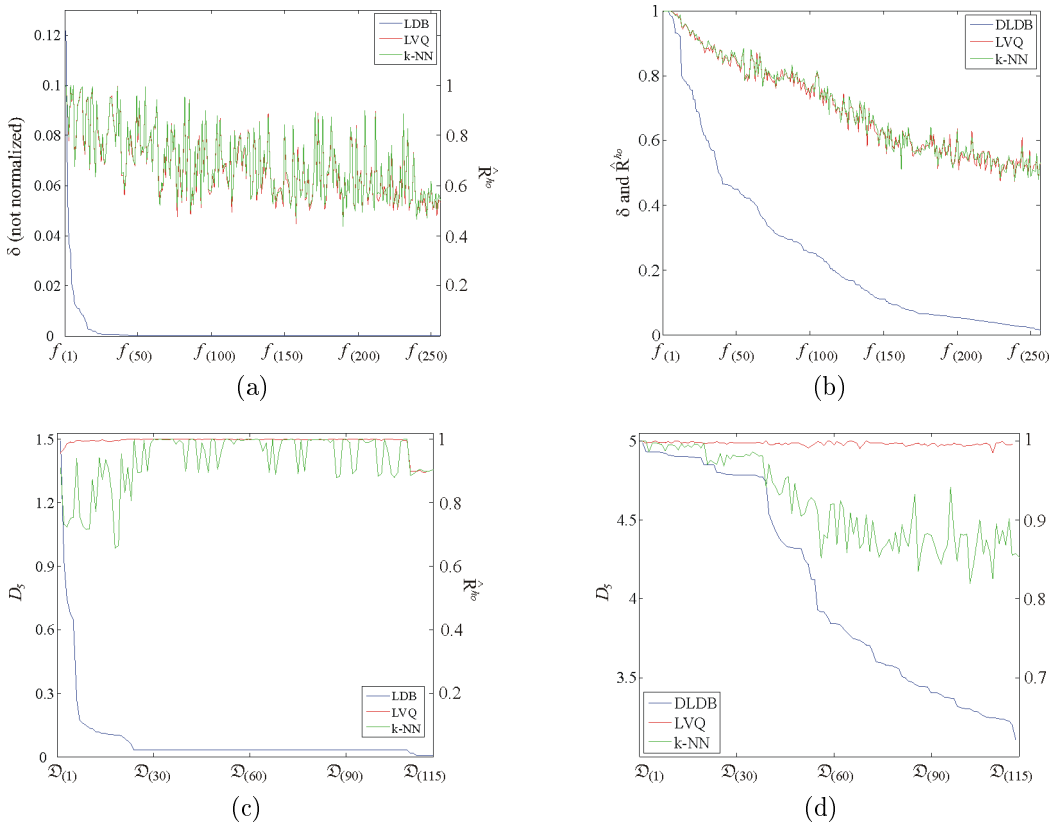


Figure 6.9: Matching the performance of LDB (left column) and DLDB (right column) with classification results. Discrimination values and identification results for both LVQ and k-NN are provided for all 256 selected coefficients of the rbior3.1 wavelet dictionary, a) and b), and all $\mathfrak{D}$ dictionaries, c) and d), by decreasing importance.

methods vastly superior.

Additionally, DLDB offered much better generalization behavior than LDB. This is proved by estimating the individual accuracies of all 256 selected coefficients and contrasting with their decreasing discrimination profile (Figures 6.9a, and 6.9b). While LDB shows a noisy profile, indicating randomness in the selection of coefficients, the discrimination profile of DLDB is consistent with our hypotheses of normalized density and discrimination-accuracy correlation proposed in chapter 5 (see section 5.4). For this two-class experiment, the accuracy of individual features $\hat{R}^{ho}$ decreases 1 to 1/2 as discrimination power reduces from 1 to 0.

The same relationship is not so evident when the length of the feature vector increases, even for the new elements next in discriminant value. When the dictionaries were compared (Figures 6.9c and 6.9d), about forty of them obtained a discrimination power of $4.75 < \delta(B\boldsymbol{x}) < 5$ with accuracies between 0.95 and 1 for the k-NN classifier. This ob-

Specie name	Country	Number of images	Fish length (cms)
M. albidus	USA	12	39-73
M. australis	Chile	20	33-109
M. bilinearis	USA	13	25-31
	Canada	18	26-50
M. capensis	Namibia	18	30-85
M. gayi	Peru	28	27-84
	Chile	25	27-108
M. merluccius	Spain (Med.)	25	25-78
	Spain (Atl.)	19	18-70
M.paradoxus	Namibia	20	27-85
M.polli	Senegal	24	34-74
M. productus	USA	21	34-74
	Canada	23	37-69
M. senegalensis	Senegal	21	35-66

Table 6.5: Available *Merluccius* material for the comparison studies developed in this chapter. Source: AFORO database (http://aforo.cmima.csic.es) and Torres (1997)

servation reinforces the previous assumption that linear classifiers, which are mainly feature independent, lose robustness with the size of the feature vector. On the other hand, classifiers of a much more complex learning stage may offer better performance in this sense for an optimum network configuration.

6.2.2 Results by Intra-specific and Inter-specific Analysis of *Merluccius* Populations

A total of 287 contours corresponding to ten *Merluccius* species from the AFORO website (Table 6.5), and a database created by Torres (1997), were extracted and normalized using the usual procedure.

The DLDB method was used to extract the relevant features and the LVQ to classify the species. However, the validation method had to be changed because the available images per species were scarce. For this reason, the *cross-validation* method ($\hat{R}^{cv}$) was chosen to estimate the identification accuracy. Calculations were developed constructing five subsets ($K = 5$) but, as in the test sample method, this calculation was repeated 50 times to ensure a value of $\hat{R}^{cv}$ with reduced variance.

In each experiment, several bin widths ($m = 3, ..., 10$) were configured and tested to

Range of δ	Expected range of $\hat{R}^{cv}$		
	2-class	3-class	C-class
0 - 0.2	[0.5,0.6)	[0.333,0.466)	$[1/C, 1/5C\,(4+C))$
0.2 - 0.4	[0.6,0.7)	[0.466,0.6)	$[1/5C\,(4+C), 1/5C\,(3+2C))$
0.4 - 0.6	[0.7,0.8)	[0.6,0.733)	$[1/5C\,(3+2C), 1/5C\,(2+3C))$
0.6 - 0.8	[0.8,0.9)	[0.733,0.866)	$[1/5C\,(2+3C), 1/5C\,(1+4C))$
0.8 - 1	[0.9,1)	[0.866,1)	$[1/5C\,(1+4C), 1)$

Table 6.6: Expected relation between normalized discrimination value and classification accuracy for C classes.

construct histogram-based densities for each class in order to reduce bias and variance errors generated by reduced data sets (Scott 1992). As the bin number could not be so large, this issue was addressed empirically by testing different widths beforehand and choosing the one with the closest discrimination-accuracy match in relation to Table 6.6. Yet, this issue will require the development of new measures or methodologies that addresses this problem more efficiently (the concluding section 7 discusses this in greater detail).

Once the bin width was set and the best dictionary was determined with the library version of DLDB, results were grouped in three categories in order to consider the previous bin width limitation:

1. *Features with* $\hat{R}^{cv} > 0.9$, or $0.8 \leq \hat{R}^{cv} \leq 0.9$, *with consistent discrimination power*[1] δ, were taken into account and their corresponding otolith irregularities were considered **relevant**.

2. *Features with* $0.8 \leq \hat{R}^{cv} \leq 0.9$ *not consistent with discrimination power* were considered **undetermined**.

3. *Features with* $0.8 < \hat{R}^{cv}$ were considered **insufficient**.

In the first group, any correct classification rate over 90% is accepted unconditionally. A little inconsistency with discrimination power may indicate that DLDB has developed a rather poor estimation of class densities. Analogously, accuracies between 80% and 90% consistent with discrimination power are very significant to be rejected, and their associated features will be taken into account.

In the second case, the DLDB has failed in the estimation of class densities, but the accuracy is so representative. At this point, the bin width limitation may be the cause

[1] With the word 'consistent' we mean that identification and discrimination are paired according to Table 6.6.

of this poor performance and the recommendation is to repeat the experiment with more samples whenever possible. If not, the risk of validating such results should be taken into account.

Finally, any result below 80% will be discarded. The observation of consistency, although strange in such identification rates, may indicate the presence of high heterogeneity within class. However, verifying this hypotheses would also require the use of a representative set of data, specially if the variances of classification accuracy are large.

6.2.2.1 Inter-specific Experimentation

Results for the inter-specific experiments in sympatric species are provided in figures 6.10 - 6.13; and tables 6.7-6.10. The information provided contains 1) the configuration settings of DWPT and DLDB: wavelet dictionaries and bin widths used in each experiment; 2) the values associated to the best wavelet atoms: the triplet $\gamma = (j, k, l)$, discrimination power vs. identification accuracy (calculated by DLDB and LVQ, respectively, so that the reader can match discrimination and classification results) and time-frequency support; and 3) the figures representing their extension of associated features on the otolith contour and their significance.

We do not provide any representation regarding the selected admissible trees of each experiment and the waveform profile of selected basis functions, as this information is represented perfectly by the dictionary name and the index triplet $\gamma = (j, k, l)$ method performance. As well as the previous *M. merluccius-G. morhua* experiment, the relevance of each feature is indicated in the figures by means of hot bars.

As expected, inter-specific results were generally more successful than intra-specific and shape differences were recognized in three out of four experiments. All rates of correct classifications were above 90% with the exception of *M. gayi* - *M. australis* (78%).

- ***M. albidus* - *M. bilinearis***

The best atoms corresponded to larger scales in the time-frequency plane (j =5, 6 and 7, see Table 6.7), of low position resolution (32, 64 and 128 contour points, respectively, see Fig. 6.10) and high frequency resolution. The most relevant feature $f_{(1)}$ is centered at the ventral side of the otolith and extended to the rostrum. $f_{(2)}$ and $f_{(3)}$ are centered along the dorso-rostral margin, and $f_{(4)}$- $f_{(5)}$ covers a small area of the rostrum (top and bottom, respectively).

All coefficients have maximum discrimination power ($\delta = 1$), allowing the correct identification of all specimens ($\hat{R}^{cv} = 1$) when classifying with $f_{(1)}$, $f_{(2)}$ or $f_{(5)}$ alone. As such,

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^{5*}$
Configuration	Dictionary: **rbior3.5**. Bin width: $\boldsymbol{m} = \mathbf{4}$.					
Atom $\psi_{j,k,l}$	$\psi_{7,0,0}$	$\psi_{6,7,3}$	$\psi_{6,1,3}$	$\psi_{5,1,7}$	$\psi_{5,1,0}$	-
$\hat{R}^{cv}$	1	1	0.962	0.957	1	1
Ac. error	0.005	0.003	0.019	0.025	0	0
δ	1	1	1	1	1	5
u	-0.621	0.571	0.57	0.23	-0.242	-
η	0.009	0.129	0.04	0.075	0.075	-

Table 6.7: Feature extraction and identification results in inter-specific *M. albidus* - *M. bilinearis*, north-west Atlantic, USA. Accuracy and discrimination values ($\hat{R}^{cv}$ and δ) are specified for each of the five basis function ψ, indexed by the triplet (j, k, l). Locating positions (u) are provided in radians, and the normalization $2\eta/N$ has been used in the discrete frequency. * The additional column shows results for the vector, $\{f_{(1)}, \ldots, f_{(5)}\}$, in $\mathbb{R}^5$.

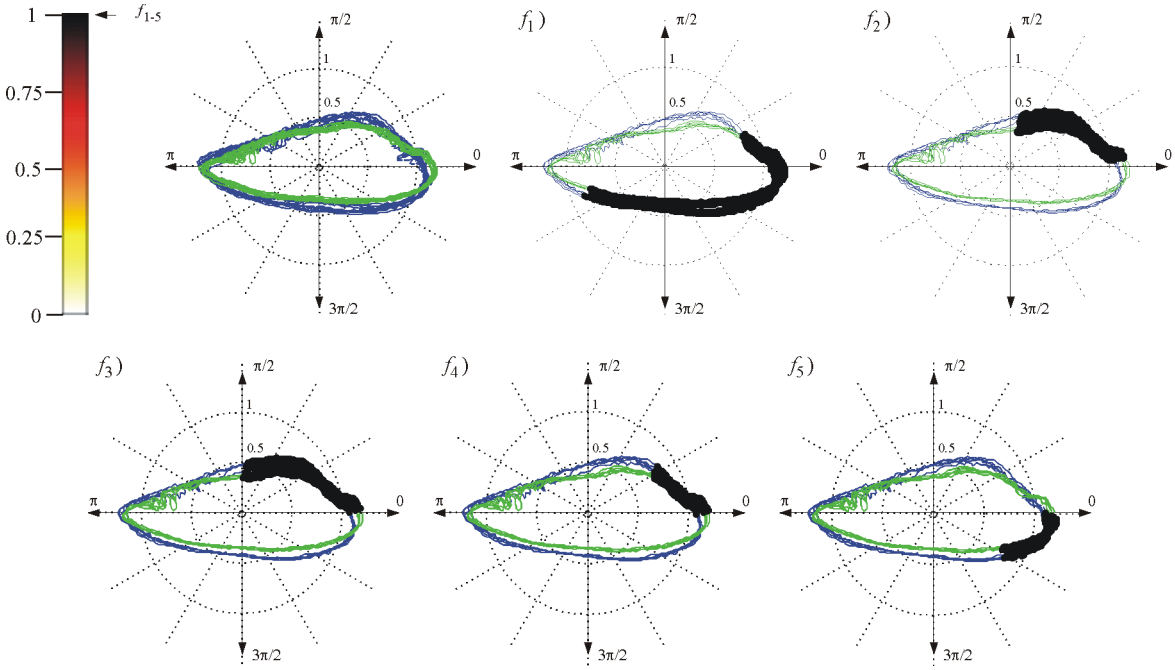


Figure 6.10: Location of otolith differences in inter-specific *M. albidus* (blue) - *M. bilinearis* (green), north-west Atlantic, USA. Feature location is expressed in terms of the momentum, u, and its variance, σ_u, over the contour trace. The colored hot bar indicates the significance of discriminant measures.

all features are considered **relevant**, concluding that the differences are extended in many parts of the contour trace, specially in the rostral part of the otolith.

These results supports the anatomical description showed by Mombeck (1970) where *M. albidus* was described as having bigger rostral development than *M. bilinearis*. The fact that almost all otolith parts had been highlighted (except near the caudal point) with maximum possible discrimination, indicates differences relative to the otolith size of both species, which is consistent with the previous contour description.

- ***M. capensis - M. paradoxus***

The selected basis functions corresponded to the largest possible scale in the admissible tree ($j = 8$) (Table 6.8). Therefore, the center position of the wavelets presented low resolution and considerable variance, more than 128 points of the contour trace for the 'rbior3.1' wavelet dictionary. While $f_{(1)}$ and $f_{(3)}$ covers the dorso-rostral and dorso-caudal margin, including the bottom part of the caudal point, $f_{(2)}$ extends its support to the left half of the ventral area (Fig. 6.11).

All five features obtained significant discriminant values ($0.69 \leq \delta \leq 0.77$) but the variance of identification accuracy was considerable compared to *M. albidus* and *M.bilinearis*. However, although the discriminant values of $f_{(3)}$ and $f_{(5)}$were not consistent, their identification accuracy was over 90%. In addition, $f_{(1)}$, $f_{(2)}$ and $f_{(4)}$ were consistent in $0.8 \leq \hat{R}^{cv} \leq 0.9$. As such, it was concluded that all five features are **relevant**. Specially, $f_{(5)}$ identified almost 93% of the *M. capensis* and *M. paradoxus* specimens.

These results are consistent with Lombarte and Fortuño (1992), and Torres *et. al* (2000b). Those studies showed a close anatomical relationship between development on the caudal part of sulcus acusticus and on the dorso-caudal margin in the otolith sagitta of both species. Since both species have common phylogeny, differences could be attributed to the adaptive characters of the inner ear and temperature effects between inhabiting the African continental shelf and deep sea water (Botha 1971). However, this hypothesis needs to be studied further.

- ***M. polli - M. senegalensis***

All coefficients obtained good classification results consistent with δ. In particular, the identification of both fish types is almost complete with $\psi_{7,67,1}$, which has the maximum possible discriminant value (Table 6.9). Additionally, none of the accuracies are below a 90%-rate, except for the feature set of five elements. As such, it was concluded that all features are **relevant** for identifying both species, where the central dorsal part between rostrum and cauda contain the most significant differences.

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^5$
Configuration	Dictionary: **rbior3.1**. Bin width: $\boldsymbol{m} = \mathbf{7}$.					
Atom $\psi_{j,k,l}$	$\psi_{8,15,0}$	$\psi_{8,23,0}$	$\psi_{8,39,0}$	$\psi_{8,135,0}$	$\psi_{8,71,0}$	-
$\hat{R}^{cv}$	0.865	0.883	0.914	0.844	0.925	0.872
Ac. error	0.042	0.047	0.031	0.052	0.031	0.039
δ	0.77	0.74	0.73	0.71	0.69	3.64
u	2.525	3.124	2.296	1.62	1.62	-
η	0.42	0.426	0.459	0.831	0.498	-

Table 6.8: Feature extraction and identification results in inter-specific *M. capensis* - *M. paradoxus*, south-east Atlantic, Namibia.

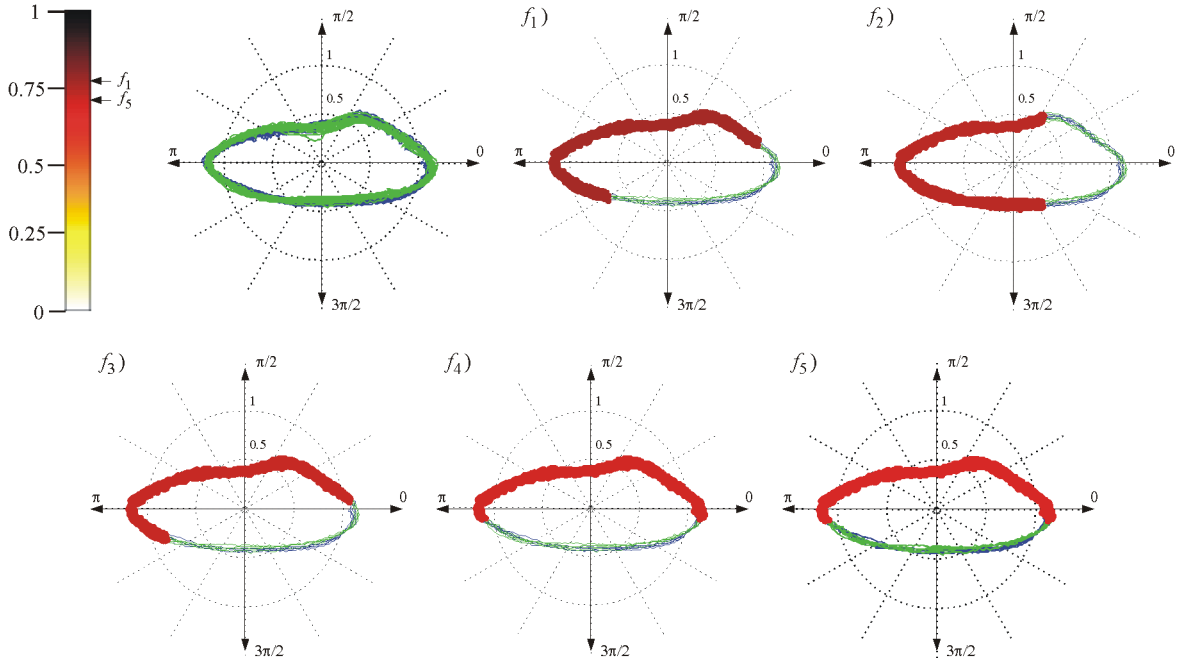


Figure 6.11: Location of otolith differences in inter-specific *M. capensis* (blue) - *M. paradoxus* (green), south-east Atlantic, Namibia.

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^5$
Configuration	Dictionary: **rbior3.1**. Bin width: $\boldsymbol{m = 6}$.					
Atom $\psi_{j,k,l}$	$\psi_{7,67,1}$	$\psi_{7,19,1}$	$\psi_{7,11,1}$	$\psi_{6,17,2}$	$\psi_{7,7,1}$	-
$\hat{R}^{cv}$	0.992	0.941	0.976	0.928	0.955	0.957
Ac. error	0.016	0.029	0.013	0.023	0.013	0.009
δ	1	0.89	0.87	0.85	0.80	4.41
u	1.606	1.707	-0.327	2.667	1.958	-
η	0.832	0.459	0.426	0.498	0.417	-

Table 6.9: Feature extraction and identification results in inter-specific *M. polli* - *M. senegalensis*, central-east Atlantic, Senegal.

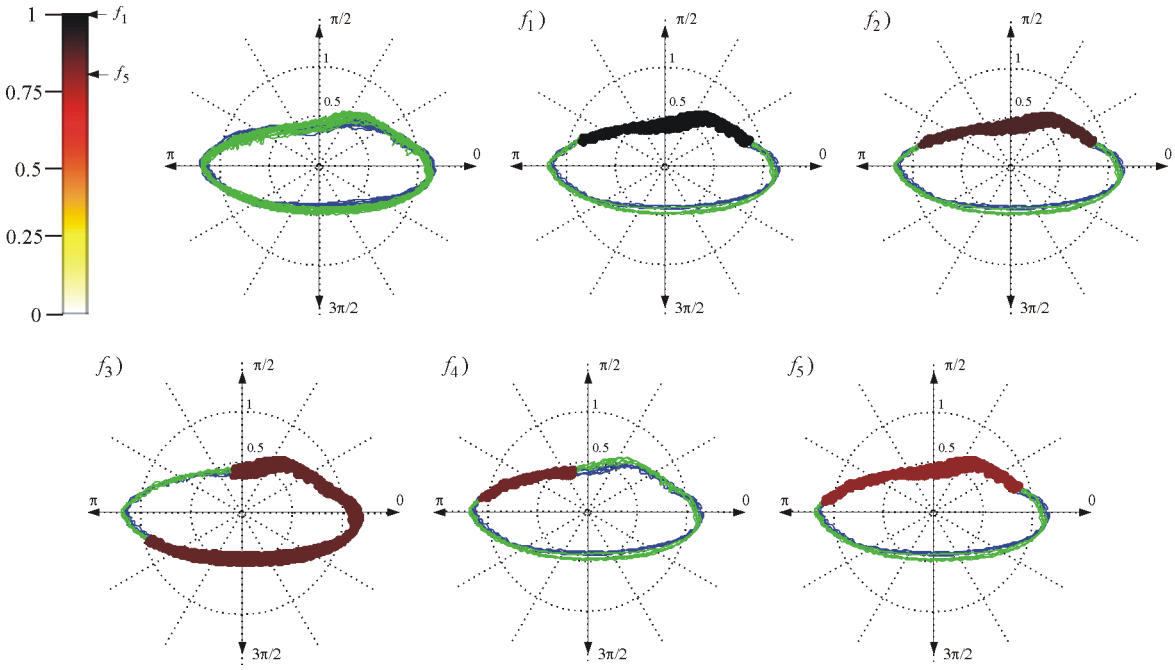


Figure 6.12: Location of otolith differences in inter-specific *M. polli* (blue)- *M. senegalensis* (green), central-east Atlantic, Senegal.

Irregularities are found at the center on the top part of the otolith ($\pi/2$) between the dorso-rostral and dorso-caudal margin with features $f_{(1)}$, $f_{(2)}$ and $f_{(5)}$, whereas $f_{(3)}$ extends from the dorso-rostral margin, through the rostral and ventral area, and $f_{(4)}$ covers the dorso-caudal margin (Fig. 6.12).

As in the *M. capensis* - *M.paradoxus* experiment, the same conclusion regarding the morphological differences and habitat could be attributed. The pair *M. polli* - *M.senegalensis* presents a depth distribution similar to these because they also belong to the African coast. In addition, this conclusion is consistent with a study developed by Roldan *et al.* (1999) who showed a similar phylogenetic relationship between both pairs. Therefore, the results of both experiments can be equivalent.

- ***M. gayi* - *M. australis***

The worst inter-specific result was obtained in the *M. gayi* - *M. australis* experiment. Feature $f_{(1)}$ is the only one surpassing the value of 0.5 which represents an identification rate of approximately 75% (Table 6.10). Its atom is well localized at the ventral side near the cauda (Figure 6.13). The use of the vector in $\mathbb{R}^5$ containing the best features did not account for a significant improvement (with an increase of only 3%). Even though all features appears to be consistent with δ, except $f_{(2)}$, these values are too small. As such, the results were considered as **insufficient**, and the conclusion is that the method failed in the identification of *M. gayi* and *M. australis* otoliths.

The cause of such a poor performance could be attributed to the high variability in the size of the species under consideration (see Table 6.5), which includes adult and juvenile individuals, as observed by Lombarte and Castellón (1991). Another reason of such heterogeneity may be the presence of species in the area which have not previously been described, such as *M. patagonicus* from Chile, which is often confused with *M. australis* or even *M. gayi* (Lloris and Matallanas 2003). If this assumption could be contrasted experimenting with supplementary and objective data, our results could serve to verify that the initial data set may be erroneous for this reason. However, this hypothesis could not be proved.

6.2.2.2 Intra-specific experimentation

Intra-specific results are provided in figures 6.14 - 6.17 and tables 6.11 - 6.14. As expected, classification values were generally inferior than inter-specific, due to the similarity of the species compared. However, some conclusions could be taken from two experiments, which may be related to geographical factors. These are the experiments corresponding to the north-west Atlantic *M. bilinearis* and south-west Pacific *M. gayi* where identification rates

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^5$
Configuration	Dictionary: **bior3.5**. Bin width: $\boldsymbol{m} = \mathbf{7}$.					
Atom $\psi_{j,k,l}$	$\psi_{1,0,50}$	$\psi_{8,140,0}$	$\psi_{6,53,1}$	$\psi_{8,141,0}$	$\psi_{5,29,7}$	-
$\hat{R}^{cv}$	0.754	0.572	0.691	0.623	0.639	0.786
Ac. error	0.056	0.078	0.058	0.085	0.075	0.056
δ	0.61	0.35	0.34	0.33	0.31	1.94
u	-2.835	3.0	-2.738	0.094	0.238	-
η	0.191	0.969	0.557	0.966	0.707	-

Table 6.10: Feature extraction and identification results in inter-specific *M. gayi* - *M. australis*, south-east Pacific.

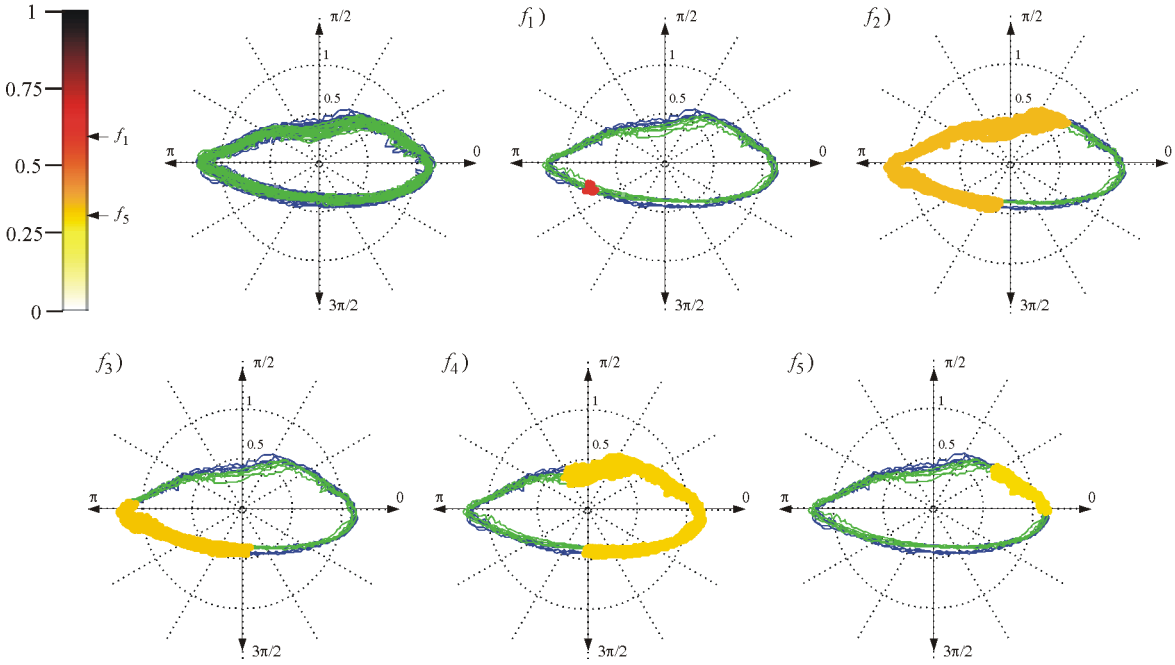


Figure 6.13: Location of otolith differences in inter-specific *M. gayi* (blue)- *M. australis* (green), south-east Pacific.

over 86% were observed.

- ***M. merluccius*: Mediterranean and north-east Atlantic (Spain)**

All discriminant values were below 0.5 (Table 6.11). The best feature $f_{(1)}$, which is positioned along the dorso-caudal margin and extends along the middle of the ventral side ($2\pi/3$, see Fig. 6.14), is consistent with δ but the accuracy of correct identifications is only about 72%. The rest of the features were not consistent: $f_{(2)}$, which has a discrimination of 0.47 is only 61.5% accurate, whereas $f_{(3)}$ which has a discrimination of only 0.31 is 71.7% accurate. The accuracy using all five coefficients is approximately 80%.

Thus, these results were considered **insufficient** and the experiment was unable to establish differences between the two Iberian *M. merluccius* populations, perhaps due to the scarce availability of samples.

A similar experiment previously implemented by Torres *et al* (2000b) obtained better results with the use of Fourier descriptors and LDA. Their work concluded that morphometric descriptors, such as size and shape, tend to be larger in Mediterranean hake because they are more sensitive to environmental factors, such as water temperature and depth. This may be an explanation for our poor results, since the size of all otoliths was normalized.

Other studies, however, suggest that genetic factors are more closely related to otolith morphology than water temperature and depth (Lombarte and Lleonart 1993, Roldan et al. 1999, Lombarte et al. 2003), which seems rather contradictory. So no conclusion could be drawn regarding the contour areas highlighted by DLDB.

- ***M. bilinearis*: north-west Atlantic (USA - Canada)**

All coefficients were consistent with the discriminant value δ but only the best feature $f_{(1)}$ was significant according to our three defined result categories (Table 6.12). The accuracies of the remaining coefficients were below 80%, and combining the coefficients from all five features did not improve correct classification rates either.

Therefore, only the shape of the basis function belonging to $f_{(1)}$ ($\psi_{8,179,0}$) which supports the dorso-rostral and dorso-caudal margins (Fig. 6.15), was considered **relevant** for distinguishing *M. bilinearis* otoliths from USA and Canada (90% of correct classifications).

These species correspond to two stocks managed separately because of differences in the exploitation patterns of both countries (Garrison and Link 2000). Although stock differences were confirmed by Bolles and Begg (2000) using the morphology of full otoliths and growth rates, no formal definition in terms of structural otolith patterns have been provided. Our

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^5$
Configuration	Dictionary: **db16**. Bin width: $\boldsymbol{m} = \mathbf{5}$.					
Atom $\psi_{j,k,l}$	$\psi_{8,27,0}$	$\psi_{4,9,7}$	$\psi_{7,12,1}$	$\psi_{2,3,32}$	$\psi_{6,17,1}$	-
$\hat{R}^{cv}$	0.721	0.615	0.717	0.615	0.601	0.795
Ac. error	0.060	0.086	0.060	0.084	0.079	0.038
δ	0.48	0.47	0.31	0.30	0.28	1.84
u	3.052	2.762	-1.878	2.656	0.209	-
η	0.068	0.905	0.065	0.62	0.492	-

Table 6.11: Feature extraction and identification results in intra-specific *M. merluccius*, Mediterranean and north-east Atlantic.

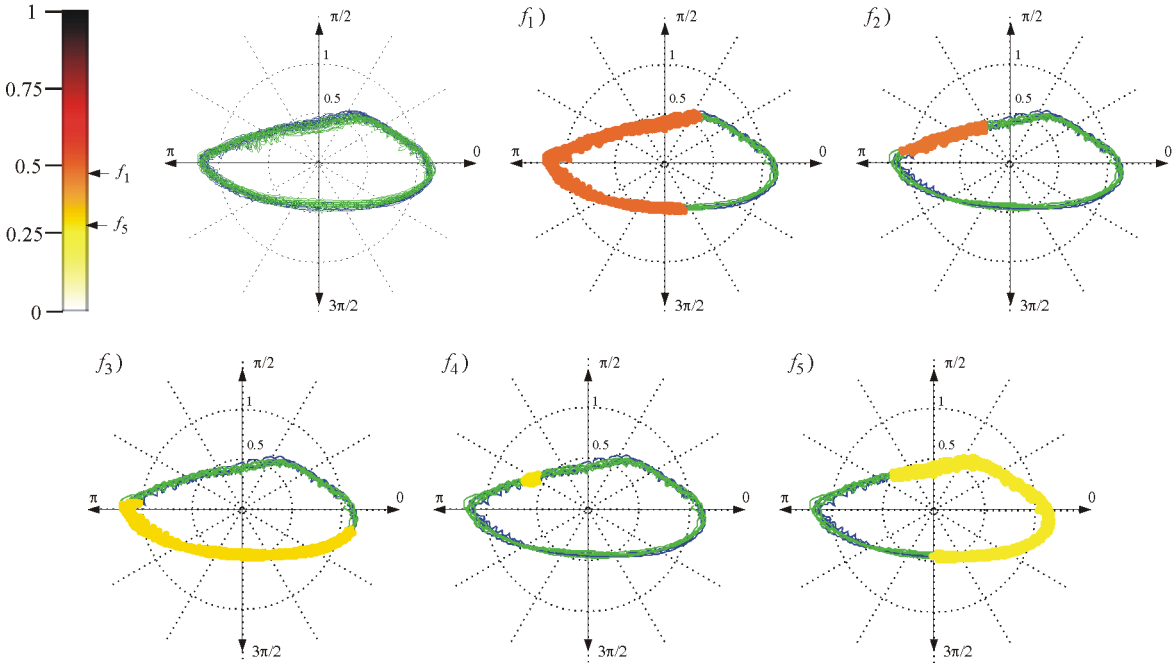


Figure 6.14: Location of otolith differences in intra-specific *M. merluccius*, Mediterranean (blue) and north-east Atlantic (green).

results indicate a high probability that these pattern differences are associated to the dorso-rostral and dorso-caudal margins.

- ***M. gayi*: south-west Pacific (Chile - Peru)**

Three features, $f_{(1)} - f_{(3)}$, presented significant accuracy (between 80% and 87%). Unfortunately, only the best one, $f_{(1)}$, was consistent with its discrimination value (0.66 against an identification probability of 86.6%, see Table 6.13). As such, this was the only **relevant**

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^5$
Configuration	Dictionary: **rbior3.9**. Bin width: $\boldsymbol{m} = \mathbf{4}$.					
Atom $\psi_{j,k,l}$	$\psi_{8,179,0}$	$\psi_{7,18,0}$	$\psi_{7,23,0}$	$\psi_{4,6,8}$	$\psi_{6,30,1}$	-
$\hat{R}^{cv}$	0.894	0.728	0.795	0.627	0.611	0.773
Ac. error	0.060	0.074	0.062	0.097	0.080	0.080
δ	0.85	0.51	0.47	0.43	0.40	2.66
u	1.973	-1.131	-1.288	3.066	-2.745	-
η	0.872	0.242	0.226	0.289	0.329	-

Table 6.12: Feature extraction and identification results in intra-specific *M. bilinearis*, north-west Atlantic, USA and Canada.

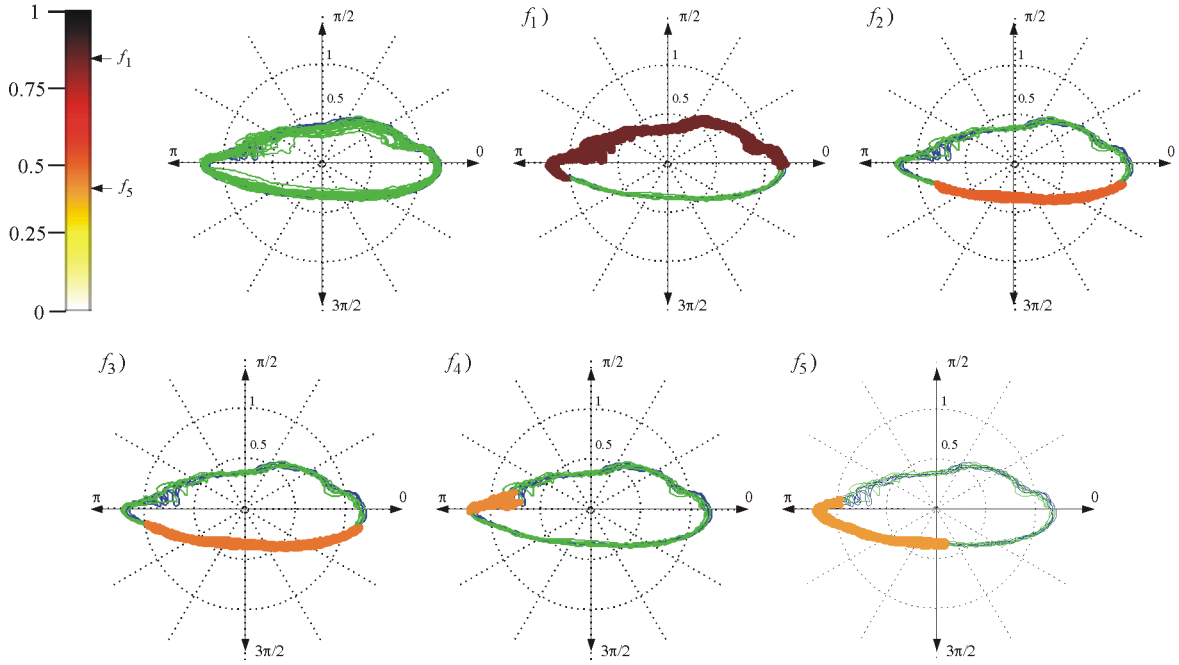


Figure 6.15: Location of otolith differences in intra-specific *M. bilinearis*, north-west Atlantic, USA (blue) and Canada (green).

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^5$
Configuration	Dictionary: **db31**. Bin width: $\boldsymbol{m = 7}$.					
Atom $\psi_{j,k,l}$	$\psi_{1,0,4}$	$\psi_{2,3,11}$	$\psi_{2,3,54}$	$\psi_{4,10,13}$	$\psi_{1,0,3}$	-
$\hat{R}^{cv}$	0.866	0.810	0.864	0.767	0.784	0.844
Ac. error	0.033	0.039	0.026	0.047	0.033	0.034
δ	0.66	0.42	0.39	0.36	0.35	2.18
u	11.69	-169.23	-22.75	178.10	13.59	-
η	31.96	79.69	79.69	99.61	31.96	-

Table 6.13: Feature extraction and identification results in intra-specific *M. gayi*, south-west Pacific, Chile and Peru.

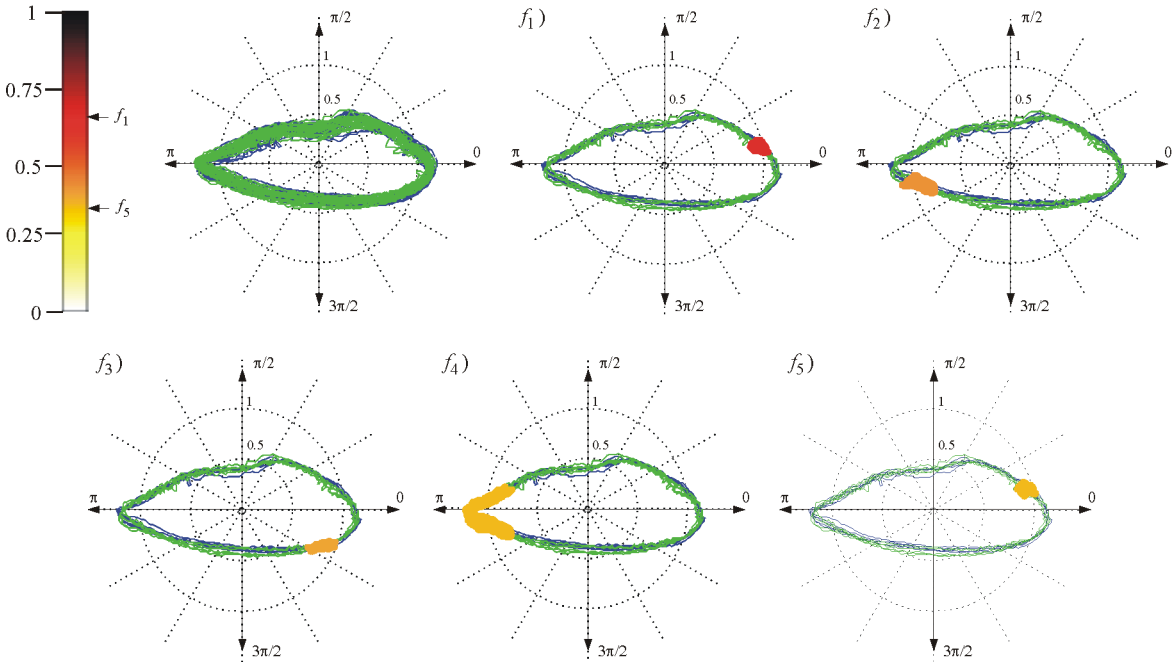


Figure 6.16: Location of otolith differences in intra-specific *M. gayi*, south-west Pacific, Chile (blue) and Peru (green).

feature for identifying *M. gayi* but we are unsure about the use that can be given to $f_{(2)}$ and $f_{(3)}$.

These three coefficients are positioned on different parts of the contour trace and have a narrower support (low scale j) compared to previous experiments: $f_{(1)}$ is localized at the dorso-rostral margin very close to the rostral area, $f_{(2)}$ is near the caudal point on the ventral side and, finally, $f_{(3)}$ is near the rostral area on the ventral side (Fig. 6.16).

This situation is analogous to the *M. merluccius* intra-specific experiment since Peruvian *M. gayi* otoliths also have relatively bigger size than those of the Chilean coast (Aguayo-

Hernández 1955). However, in this case a very local relevant feature, a notch in the dorso-rostral margin (Fig. 1.4), is significant between both fish groups. This observation could confirm the influence of factors, such as water temperature, narrowing of the continental shelf in the south-western Pacific or the presence of hydrogen sulphide in deep water layers between latitudes 13th and 19th, as concluded from Torres *et al* (2000b).

- ***M. productus*: north-west Pacific (USA - Canada)**

The lowest discrimination results of the intra-specific experiments were obtained for the *M. productus* specimens. The discrimination corresponding to the best feature was too small (only 0.5) and none of them were consistent with identification accuracy, except for $f_{(5)}$ (Table 6.14).

$f_{(1)}$ and $f_{(2)}$ are positioned on the rostral area, $f_{(3)}$ and $f_{(4)}$ on the dorso-caudal margin, and $f_{(5)}$ at the caudal point. All accuracies were below an 80% rate, including the combined feature vector in $\mathbb{R}^5$ that did not not even reach the value of 70% (Fig. 6.17). As such, all the features were **insufficient** and DLDB was unable to classify *M. productus* specimens.

Recent genetic studies in these populations have showed a high level of similarity, possibly related to the inhabiting area they share, the north-west Pacific (Silva-Segundo et al. 2011). Although more studies should be developed to definitely confirm this hypothesis, our results would indicate that the specimens of both geographical locations are actually the same.

6.2.3 Discussion of Results

6.2.3.1 Feature Selection Methodology

The DLDB methodology has been tested and compared with other analytical tools in fish identification experiments. This new methodology is based on the Best-Basis paradigm. Although the comparative *M.merluccius-G. morhua* study is not of interest at all from the perspective of ichthyology, its results allowed the DLDB to be compared with LDB and other standard tools, such as LDA, PCA or DFT.

Three issues have been addressed with regard to the Best-Basis paradigm. Firstly, DWPT coefficients have been evaluated by estimating their distribution of density, instead of the energy magnitude of the original proposal. This modification allows the coefficients with greater separation characteristics to be selected and the performance of classifiers to be enhanced. In our opinion, this measure is also appropriate for coefficients of smaller energy, which are also useful for classification if their class separability shows significative.

Features	$f_{(1)}$	$f_{(2)}$	$f_{(3)}$	$f_{(4)}$	$f_{(5)}$	$\mathbb{R}^5$
Configuration	Dictionary: **rbior3.7**. Bin width: $\boldsymbol{m = 6}$.					
Atom $\psi_{j,k,l}$	$\psi_{3,2,31}$	$\psi_{3,3,31}$	$\psi_{5,0,5}$	$\psi_{4,13,10}$	$\psi_{5,20,3}$	-
$\hat{R}^{cv}$	0.766	0.705	0.568	0.707	0.660	0.690
Ac. Error	0.064	0.054	0.069	0.057	0.062	0.048
δ	0.37	0.36	0.34	0.33	0.33	1.73
u	3.86	3.86	133.60	146.36	-170.05	-
η	59.03	42.14	3.07	75.58	98.87	-

Table 6.14: Feature extraction and identification results in intra-specific *M. productus*, north-west Pacific, USA and Canada.

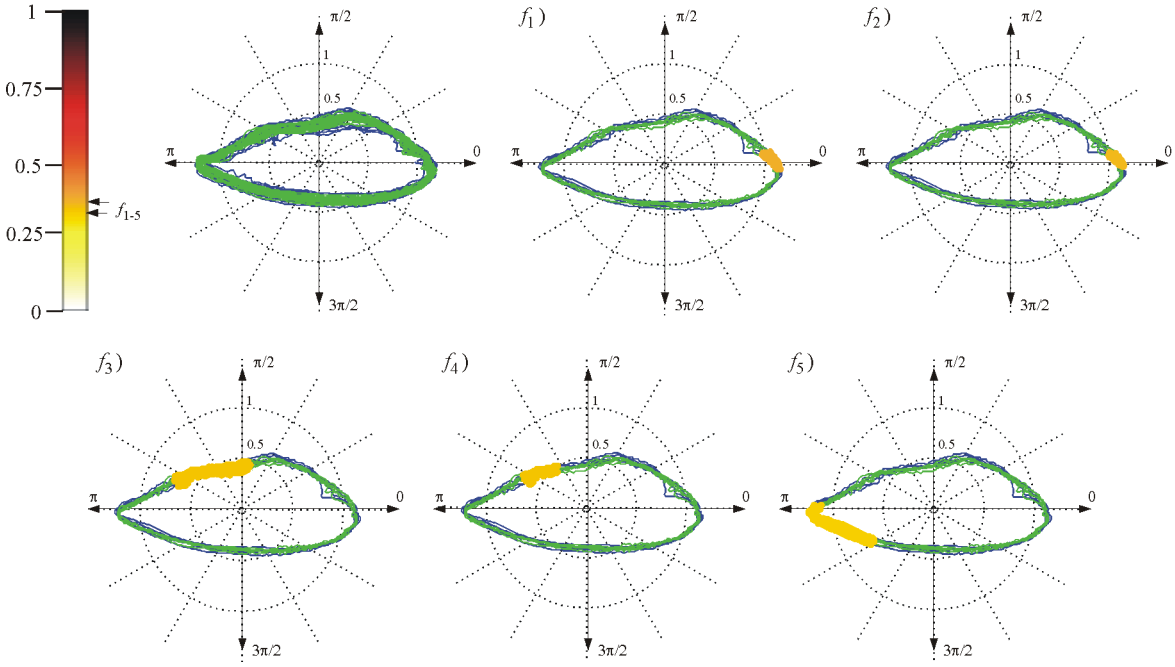


Figure 6.17: Location of otolith differences in intra-specific *M. productus*, north-west Pacific, USA (blue) and Canada (green).

Secondly, the original bottom-up strategy has been replaced by a top-down algorithm that selects DWPT nodes on the basis of their most significant element. This was modified to permit the consideration of these coefficients for classification.

And finally, the Kullback-Liebler metric was changed for a version supported on the Battacharyaa affinity measure (Eq. 5.24). The new measure has its own mechanism for representing the lowest and highest discrimination capacity, a behavior necessary for achieving data generalization and indicating the significance of structure irregularities on the image.

Our proposal can provide further information on the physical structures that make up the signals. Not only did the selected features obtained higher classification accuracy, but they also explained shape differences in terms of geometrical characters of the contour, such as the position of oscillating of structures, more closely related to ecological and biological aspects of fish species. Inferring such kind of conclusions in methods such as PCA or LDA is impossible since the eigenvectors, determined from between-class and within-class covariance matrices, are difficult to interpret from a physical point of view.

The cod-hake experiment showed that contour information is sufficient for fully identifying both fish species. The fact that STD can be even more effective than DFT (since only one PCA coefficient was necessary to obtain full identification with LVQ, whereas DFT needed of five coefficients), suggests certain conflict between PCA and DFT methods.

In fact, the implementation of PCA on DFT does not make much sense from the perspective of signal processing, as PCA also develops its own orthogonal projection, similar to DFT, for representing the components of large variance. In LDA, however, the same is not true because orthogonality is lost, making the output features very effective for classification purposes.

Analogously, these results also showed that compression and discrimination tools do not necessarily have to be considered for different purposes, since full identification is also possible for all PCA, LDA, LDB, and DLDB. From the author's point of view, this issue is conditioned by the true nature of input data. In the cod-hake experiment, both classes are teleostean species from species of different taxonomy, so it would be reasonable to expect that the covariance matrix fulfill the same discrimination role.

Our identification results also showed differences between the performance of linear and nonlinear classifiers. While LVQ proved to be more robust at increasing the input vector with more features, the accuracy of *k-NN* decreased drastically. This is a typical performance in nonlinear, which base their operation on the separation of class decision boundaries and tend to produce more accurate density estimates than Bayesian methods. In contrast, Bayesian methods are configured easily, have low computational cost and serve to obtain a first reference estimate.

Finally, another important issue concerns the proposal of new discriminant measures to represent the accuracy of correct classifications. Such measures should also be able to provide generalization regardless of the type of application or the data under study. The author sustains that this can only be achieved by ensuring the two hypotheses established in section 5.4. This relationship has been demonstrated in this experiment representing the discrimination profile of selected coefficients in the rbior3.1 dictionary. Standard methodology such as LDB, PCA and LDA among others, do not possess this property.

6.2.3.2 Ecological Significance of *Merluccius* Populations

Different *Merluccius* species have also been compared by means of the proposed DLDB. Three in four of our inter-specific experiments and two in four intra-specific comparisons succeeded in locating contour irregularities with high identification accuracies.

The inter-specific *M. albidus* and *M. bilinearis* were fully recognized, the accuracy of *M. polli* - *M. senegalensis* was 99% and *M. capensis* - *M. paradoxus* was almost 90%. The accuracy of the intra-specific *M. bilinearis* was close to 90%; and in *M. gayi* this value was over 86%.

Some results were those that would be expected based on previous studies. An earlier study developed by Mombeck (1970) revealed that *M. albidus* otoliths are wider than *M. bilinearis*. This conclusion is consistent with DLDB results, since otolith width could be viewed as a morphological parameter associated to many parts of the contour trace, except for the dorso-caudal margin and caudal point.

In the *M. capensis* - *M. paradoxus* experiment it was very difficult to associate the selected contour parts to the morphological differences with respect to otolith average length, height and total fish length as concluded by Botha (1971). However, the irregularities associated to the dorsal margin may be attributed to a difference in development of the sulcus acusticus of deep sea water fish, where climatic conditions are quite different than those of the African continental shelf (Lombarte and Castellón 1991, Torres et al. 2000b).

Results from the *M. polli* and *M. senegalensis* specimens were parallel to the *M. capensis* and *M. paradoxus* experiment, with the main differences detected in between the dorso-caudal and dorso-rostral margins (6.11 and 6.12). These differences could be associated to the development of parts from the *sulcus acusticus*, since *M. paradoxus* and *M. polli* have been described as being larger in size in order to stand the habitat conditions of deep sea waters (Inada 1981, Lombarte and Fortuño 1992, Lloris et al. 2005).

In the intra-specific north-west Atlantic *M. bilinearis*, stock differences were detected along the dorso-caudal and dorso-rostral margins by means of feature $f_{(1)}$ with an accuracy

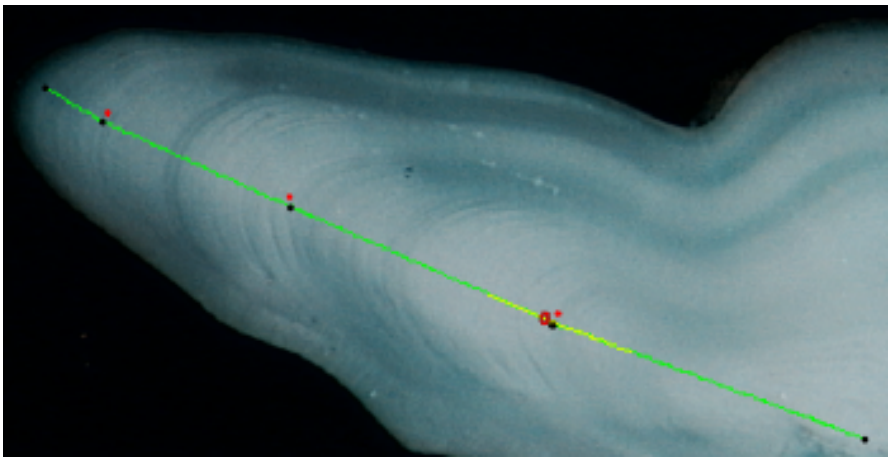


Figure 6.18: Example of automatic detection of the first year mark by means of AGD. The otolith section correspond to a 3 year-old cod specimen. The one-dimensional intensity signal has been extracted from the nucleus to the contour border (right to left) following the green trace. Age information is provided by the expert (red dots) in order to compare the detection of the first year period (red circle).

around 90% (Fig. 6.15), whereas in intra-specific north-west Pacific *M. gayi* differences could be associated to changes in temperature conditions of both stocks (Torres et al. 2000a).

DLDB was unsuccessful with inter-specific *M. gayi* - *M. australis*, intra-specific *M. merluccius* and *M. productus*. In the latter, results may indicate that both otoliths from the Pacific in Canada are, in fact, from the same population. On the other hand, the consistency with δ found in four selected features in *M. gayi* - *M. australis* may indicate that some kind of specific variability exists within the same species. However, other tests of similarity should be carried out to confirm these hypothesis, since specimens from *M. patagonicus*, whose otoliths have never been fully described, have been recently found inhabiting the same area.

6.3 Application by Age Estimation

In order to illustrate the performance of AGD, we first show the results for an image section corresponding to a three-year-old cod otolith (Fig. 6.18) . Figure 6.19 represents its extracted intensity profile $i\,[n]$ along the main growth radial axis (green line in Fig. 6.18). The signal was resampled to N=512 points and the radial length was normalized in the range $[0, 1)$. Information related to the position of year marks was also provided by the expert (red lines).

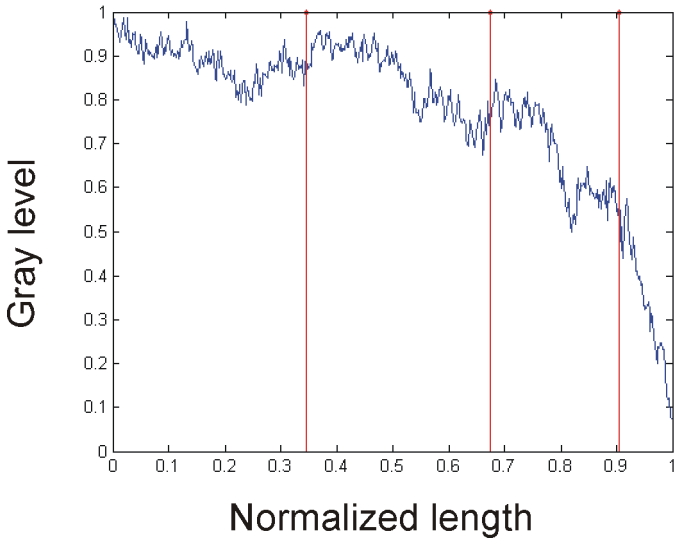


Figure 6.19: Extracted intensity profile from the previous cod otolith section in Fig. 6.18. As indicated by the expert (see red marks), year increments don't have a regular rate because of the non-linear fish growth.

6.3.1 Manual vs. Automatic Contrast Cancellation and Signal Demodulation

Both manual and automatic demodulation methods (SGD and AGD) were configured with the same VBGF parameters, $L_\infty = 1.743$ centimeters and a Brody growth coefficient $k = 0.237$, respectively, but the AGD algorithm used a window of length $W = 64$, a threshold $\gamma = 0.8$ and 32 bins for the histogram. Figure 6.20 shows the demodulation results that were obtained after estimating the optimal growth law for these methods. For the AGD method, the estimated periodicity of age structures computed for this specimen presented a normalized length, $u = 0.293$. This result, was obtained after considering the center position of the histogram bin with the most lags. Although the signals computed from both demodulation methods presented deviations, the first year mark position computed with AGD was practically the same as that of the expert in the otolith image (note the red circle in Fig. 6.18).

To study the contribution of each preprocessing method to the correct estimation of age, an experiment has been set up taking the DLDB and the LVQ classifier as the main feature extraction tools of our pattern recognition system. The experiment was carried out in the context of the AFISA European project and consisted in the estimation of age from 189 plaice specimens using otolith images provided by the Marine Research Institute of Iceland

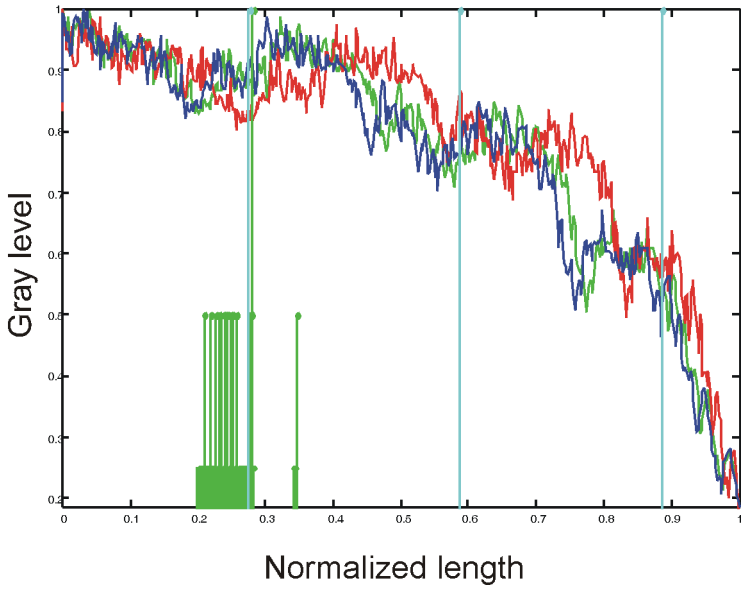


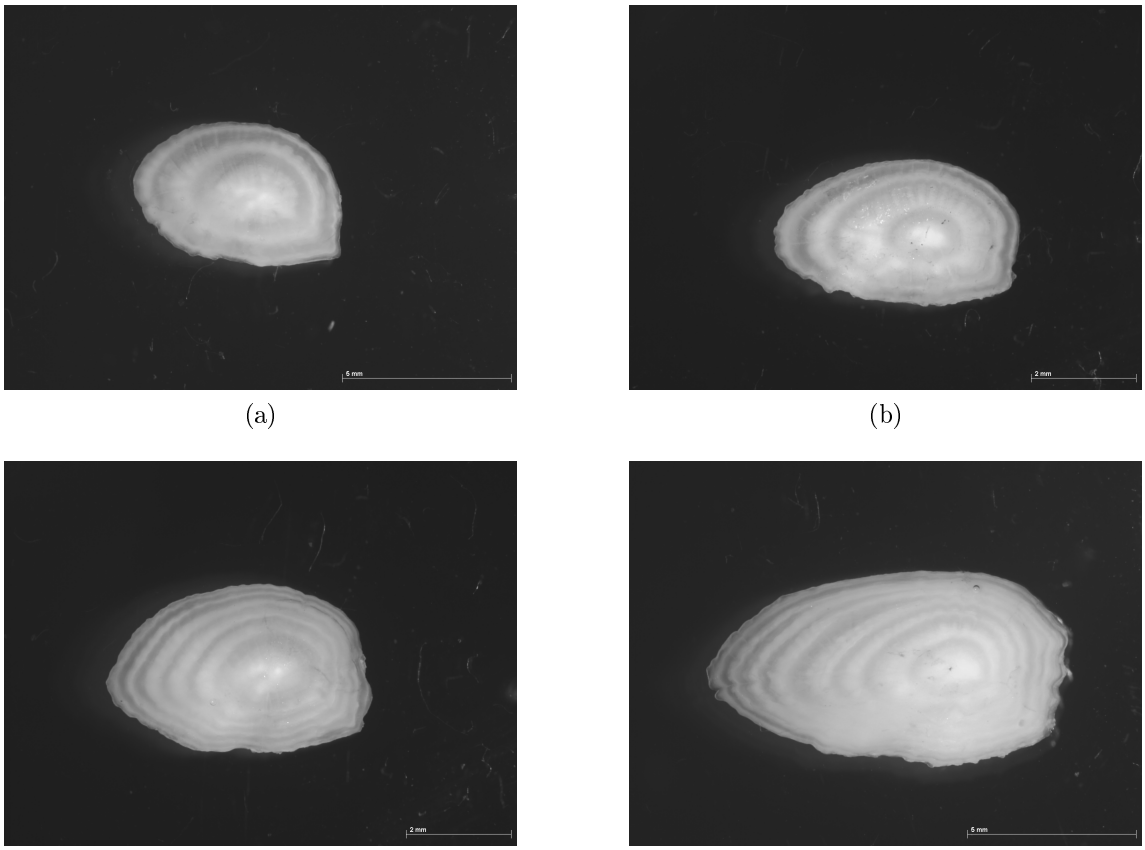
Figure 6.20: Supervised vs. Automatic Growth Demodulation results. Original, manual and automatically demodulated signals (red, blue and green traces, respectively) are represented together with the histogram of autocorrelation lags (green stems). The bin with the majority of elements sets the first year period in the demodulated signal (first blue mark) and establishes an annual rate reference for the remaining year marks.

(MRI), see Fig. 6.21. The true age of fishes were known and grouped in the following classes: **1)** up to two years old (63 otoliths); **2)** three years old (54 otoliths); **3)** four years old (23 otoliths); **4)** five years old (36 otoliths) and; **5)** six years old, or more (13 otoliths).

All the information relative to the otoliths, including gray-scale values of the intensity profiles and coordinate points corresponding to year marks and nucleus, was provided by the IFREMER institute who used the TNPC software to process the reflected-light images and to extract the one-dimensional intensity profile. Thus, year marks t_k, length values, l_n, and signal vectors $i\,[n]$ (resampled to N=512 points) were extracted from this data.

Then, the four preprocessing methods considered in this study (the MCE, AGF, SGD and AGD) were configured. For the MCE method, the detrended signals $\boldsymbol{i_C}$ were obtained subtracting the contrast level $\boldsymbol{i}_D$ from the original profile, $\boldsymbol{i}$. $\boldsymbol{i}_D$ was set thanks to the interpolation of the $\{c_k\}$ contrast reference points, which were set manually after careful examination of each image. As for the AGF method, the width of the quantization filter was set to $\sigma_Q = 16$ samples and $\sigma_C = 200$ samples for the contrast filter, whereas SGD and AGD were configured with the same initial parameters as the previous experiment.

To contrast the four preprocessing methods, the DLDB was used to obtain a first estimate of correct age classifications with the original intensity profiles. The library version of the



(a) (b) (c) (d)

Figure 6.21: Plaice otolith samples used in our experiments for different age classes. (a) two years; (b) three years; (c) five years; (d) six years or more.

DLDB evaluated the best type of standard basis function for a decomposition depth of J=9 levels in the DWPT tree, using 9-bin length histograms to determine the density distribution of the coefficients. The feature vector containing the five best discriminant coefficients, $\{\boldsymbol{w}_{(n)}\boldsymbol{x}\} \in \mathbb{R}^n$ for n=5, was introduced as input data to the LVQ classifier to estimate the age of all specimens. An average rate of correct classification of **55%** of the samples, $\hat{R}^{cv}(r) = 0.55$, was obtained after 50 trials with the 'k-fold cross-validation' method, using K=5 groups with $N_s \simeq N/K = 37$ samples in each group.

The preceding configuration was then used to evaluate the efficiency of the four pre-processing methods, first independently (Table 6.15), and then combining together contrast cancellation and growth demodulation in their two modalities: manual and automated (Table. 6.16). All aging rates were estimated for the 'symlet13' waletet packet dictionary, which obtained the highest result (**88.5%**) when MCE and SGD were combined and the DLDB was computed.

	Manual	Automatic
	MCE / SGD	AGF / AGD
Contrast Cancellation	70%	70%
Growth Demodulation	81.4%	47%

Table 6.15: Aging precisions values combining DLDB with the four preprocessing methods: MCE, SGD, AGF, AGD. All four methods are considered independently and values are expressed in terms of agreement to expert estimations.

	Contrast Cancellation			
	MCE (Manual)		AGF (Automatic)	
Growth Demodulation	$\hat{R}^{cv}(r)$	$\hat{R}^{*}$	$\hat{R}^{cv}(r)$	$\hat{R}^{*}$
SGD (Manual)	88.5	75.3-84.1	85	72.25-81
AGD (Automatic)	47	40-44.7	–	–

Table 6.16: Fish aging results combining contrast cancellation and growth demodulation. Classification results are provided in terms of expert readings and their known age precision error, $\hat{R}^{*} = APE \times \hat{R}^{cv}(r)$, in terms of expert agreement which is situated between 85% and 95% for plaice otoliths.

Removing only the image contrast improved classification accuracy up to **70%** and no significant differences between manual and automatic behavior were observed. The only preprocessing method exceeding these results was the manual growth demodulation method (SGD), with a correct classification rate of **81.4%** in relation to expert estimations. Results using the AGD method were rather poor (**47%**).

When contrast cancellation and growth demodulation were used together the rate of correct estimations increased up to **88.5%** in the case of manual implementation (MCE and SGD), and **85%** when using automatic contrast cancellation (AGF and SGD). These rates are similar to those obtained by Fablet and Josse (2005). However, the poor results obtained by means of the automatic demodulation method seem to indicate that an optimal demodulation of age structures may be associated to more factors than those we have considered in the AGD method.

6.3.2 Discussion

Our results show that applying feature extraction alone to the initial intensity profiles, without preprocessing growth demodulation and eliminating image contrast, is too simple to provide relevant clues to infer fish age. These operations not only are necessary but also may not be enough for the considered easy-to-read fish species. So more expert knowledge

is required to address this problem.

From the two preprocessing methods used, growth demodulation seems to have more impact on the correct estimation of fish age than contrast cancellation, making fish growth the most decisive factor when estimating fish age. Nevertheless, the best results are obtained when both techniques are combined. This is clear from the fact that manual demodulation (SGD) and contrast cancellation (MCE) are only 81.4% and 70% accurate, respectively, whereas when both methods are combined a rate of almost 90% is obtained.

Results also show that automatic methods are far less accurate than the manual methods. The performance of automatic contrast cancellation (AGF) seems to be similar to that of its manual version whereas some conflict appears to be present with automatic grow demodulation (AGD) when used either independently or in combination with manual contrast cancellation. Its results, which were only 47% accurate, conditioned our experiments and both automatic contrast cancellation and demodulation were not considered at all. In our opinion, a possible explanation is the type of noise that could be associated to the intensity profile, which may be much more complex than initially expected. The discussion of this topic will be extended in the next chapter and future developments will be proposed.

6.4 Conclusions

6.4.1 Feature Extraction

Different *Merluccius* species and populations have been compared by means of the proposed DLDB, whose discrimination criterion is in our opinion more robust at detecting class irregularities in signals than that of LDB. We believe this contribution unifies and addresses the development of more generalized discrimination measures.

It has been mentioned that differences were detected in the otolith contour radials of inter-specific *M. albidus* - *M.bilinearis*, and higher accuracies were observed in *M. polli* - *M. senegalensis*, *M. capensis* - *M. paradoxus* and intra-specific north-west Atlantic *M. bilinearis*. However, similar results could not be obtained between Mediterranean and north-east Atlantic *M. merluccius*, south-west Pacific *M. gayi* and *M. australis*, and north-west Pacific *M. productus*.

We suspect the unsuccessful classification is due to the lack of samples, which may have influenced the precision of density estimation. This problem generated considerable bias and variance in the estimation of their corresponding densities as well as in the validation of accuracy, which produced over-optimistic results.

There are always limitations to every algorithm and of course DLDB is no exception.

There are four main concerns which we are currently investigating. In particular, four issues should be addressed in the future. The most important one focuses on the way of calculating class densities, since estimating the precise true distributions is crucial in the automatic selection of DWPT coefficients. For now, we used histograms because they are the simplest form and the first natural choice of non-parametric estimation, although they do present some drawbacks.

On one hand, the discontinuities of the estimates are not due to the underlying density, but to an artifact of the chosen bin locations delimited by the starting position (t_0) and the bin width (h). Its density estimate, $f = \frac{v_m}{hN}$, becomes more accurate as we increase N and reduce h. However, since N is fixed, to improve accuracy we could let h approach to zero but then the region $G_m \cong t_m - t_{m-1}$ would become so small that it would yield no samples ($v_m = 0$). In practice, that means we need to find a compromise for N and h, which should be large enough to include sufficient samples but also small enough to assume that $\boldsymbol{f}$ is continuous within G_m.

The use of alternative methods, such as Kernel Density Estimation (KDE) or k-Nearest Neighbor, and their capacity to improve this stage of the Best-Basis paradigm is discussed in the next chapter.

Another limitation is associated to the *translation dependence* of the DWPT method. This occurs because the sub-sampling operation destroys translation invariance of the wavelet transform, which means the control over homology of geometrical shapes is lost.

Our way to tackle this problem was to use a normalization method based on the Fourier ellipse, which modified otolith size, sample position and orientation. However, addressing this problem from other perspectives may be more effective for other types of signals. For instance, the use of customized filter-banks or the *dyadic* version of the wavelet transform, the *algorithm à trous*, which does not use sub-sampling.

The third limitation is inherent to the selection of nodes and its *search strategy*. The admissible tree varies significantly with the number of signals N or the bin width configuration (h). This issue may be closely related with accuracy validation methods, such as hold-out or cross-validation. In this sense, the introduction of such validation techniques within DLDB could allow to develop a more robust selection based also on the variances of discriminant measures, or to develop a new and completely different classifier, if a class decision rule is included at the end of the process.

Another issue is related to the property of exact reconstruction. In this sense, one may argue whether exact reconstruction is really necessary in classification tasks. Our opinion is that relaxing such condition may lead to more efficient and faster search strategies, since comparing parent and child nodes may no longer be necessary. This would allow density

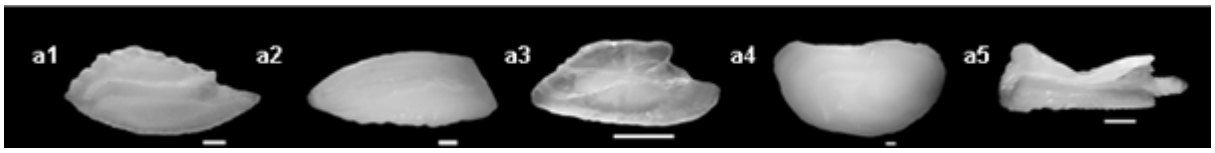


Figure 6.22: Some otolith examples from different taxonomic value. a1) *Trachurus mediterraneus;* a2) *Trisopterus minutus;* a3) *Coris julis*; a4) *Umbrina canariensis* and a5) *Scomber colias.*

functions to be constructed more easily compared to those of full nodes, where histograms of higher dimension would be necessary.

The automatic selection of features deals with other pattern recognition topics, such as the minimum description length (the optimal size of the feature vector in $\mathbb{R}^k$) and the classification of several group classes ($C>2$), among others. In the current approach, only one-dimensional signals were considered and feature vectors of five elements were used. The class densities were computed individually for each of the coefficients, which were ranked and selected according to class pairwise discriminant measures.

Not all the topics could be addressed in this work, even though they deserve similar attention. These are not associated to the DLDB mechanisms but to other more external issues, such as the necessary dimension of initial data for correct contour representation or the type of preprocessing necessary for the analysis of its structures. Some otoliths, for example, have complex concavities (Fig. 6.22) and the use of single radials can alter considerably the original representation, making it unsuitable for feature extraction purposes. This approach would require an elliptic framework which contemplated both x and y axes. The processes associated to DLDB will then need to be extended for compliance with bi-variate data, two-dimensional images or three-dimensional shapes.

6.4.2 Signal Demodulation

DLDB was also tested in the estimation of age of plaice otoliths by means of gray-scale intensity images. The manual preprocessing methods demonstrated the importance of removing image contrast variation and demodulating growth variability in order to succeed in this goal. Our results, however, showed that the performance of a computer-based and fully-automated aging method based on reflected-light images is, for now, far from being as accurate as the manual methods.

Despite these results, we believe our AGD method lays the foundations for more deterministic approaches of fish aging than current statistical approaches. Instead of developing an average growth model that is used for all fish, the AGD tries to determine the specific de-

viations of periodicity of a single otolith along its intensity profile, and uses this information to find its optimal growth function. In our opinion, this approach is necessary if we are to develop more specific computer-based aging systems that can cope with the more intrinsic and complex factors involving the growth variability of fishes.

Although manual methods seem to indicate that this growth variability may be the most influential factor, the poor results obtained by AGD suggest that other unconsidered sources of noise may be present. These sources will be investigated in the future. In this sense, the decreasing trending profile observed along the intensity signal may be closely related with image light conditions, since light is known to have a diffused reflection which goes in many different directions, specially near the otolith border whose surface is very concave. In our opinion, this is a problem closely related to the geometrical laws the entire acquisition system, including otolith and camera, are subjected to.

So solutions resort to operations at the mechanical level of image acquisition. For example, the development of an image acquisition interpolating several shots with different light settings may help to reduce (or eliminate) the high contrast variability near the contour borders and, subsequently, the steep slope at the right end of the intensity signal.

Our current design develops an implementation which combines both filter-based and feature-based techniques. Future automated methods, however, are aimed at using only filter-based techniques to estimate the exact fish age, something feasible if age structures can be demodulated so as to appear periodically in the in tensity profile. We will come back to this issue in the final chapter.

Chapter 7

Final Remarks and Further Development

7.1 General Conclusions

The problem of automatic detection and extraction of structure patterns from one-dimensional signals has been addressed in this thesis. Our work, which is based on the Best-Basis paradigm, was developed parallel to recent LDB proposals that do not address the problem of data generalization and measure standardization (Marchand and Saito 2012). These limitations have been addressed with DLDB, which has been used in the field otolith-based fish identification and aging.

In general, PCA and LDA methods do not attempt to interpret feature patterns in signals. The scatter plots only provide visual information about the separability of features, and eigenvalues are not qualitative information measures for developing automatic feature selection.

On the other hand, the efficiency of many descriptive tools for representing structural components in signals is very constrained. Although DFT is usually a good tool to describe frequency and phase information, it is in fact unable to capture localized structures (edges, spikes and ridges...) since the time domain is lost.

LDB tries to enhance classification by using wavelet packet analysis (Coifman and Wickerhauser 1992, Saito 1994, Wickerhauser 1995) but its lack of generalization has been demonstrated in this work. Its separability measures confuse energy distribution with energy magnitude, which means useless coefficients may be selected within the process, while useful ones are discarded.

Our proposition was to exploit nonparametric density approaches and the use of bounded measures, as suggested by Saito (Saito et al. 2002). With these ideas a new LDB algorithm, the DLDB (Fig. 5.2), has been implemented. This scheme uses a new top-down strategy for

wavelet trees, which prioritizes the relevance of single coefficients in the selection of nodes. The new tool provided results consistent with some classifiers, such as the k-NN and LVQ.

The DLDB was applied in fish identification and fish aging, two important aspects of the fish industry. All the experiments were implemented in the context of the AFORO3D project and the AFISA project from the European Union.

In the first application, a preprocessing method for the normalization of the otolith contour was also developed. This step, which develops automatically, minimizes the contour translation invariance of all otoliths by normalizing their size, rotation and translation. The whole scheme was used to carry out two experiments, one to classify cod and hake species to show the functionality of our proposition, and another to study the inter-specific and intra-specific relationships of *Merluccius* fish species.

In the first experiment, irregularities were found at the dorso-rostral and caudal margins of both species, obtaining the highest possible discrimination rate in relation to the proposed measure. This result was consistent with classification results, as all the specimens of both species were fully identified (Table 6.6).

In the study of hake species, the best inter-specific identification rate corresponded to the *M. albidus* - *M. bilinearis*, followed by *M. polli* - *M. senegalensis*, and *M. capensis* - *M. paradoxus*; whereas the best intra-specific results were obtained for *M. bilinearis* and *M. gayi*.

As for the field of fish aging, a very novel approach for automatic demodulation of growth structures for reflected-light images, the AGD, has been proposed. The method introduces a new philosophy based on adapting *template functions* to the specific growth characteristics of individual fish. Although results were unsuccessful, this new philosophy represents a totally different approach of automatic aging compared with classical statistical methodology and should lay the principles for single fish growth estimation.

DLDB is a reference tool towards the full automation of signal detection. Not only it is able to find class patterns in signals but also explains them in terms of physical composition. AGD, on the other hand, represents a new philosophy for fish growth estimation. Although their use is restricted here to applications within the fish industry, they can be exploited in other fields, such as the analysis of electrical signals, medical applications or geophysical phenomena, among others.

In this sense, we believe they offer scope for range of promising computer tools in signal processing and pattern recognition, to be explored in the near future. The following sections outline some of the current problems of both DLDB and AGD, and pose new challenges.

7.2 Feature Extraction Methodology

7.2.1 Input Data

As all our proposals work with one-dimensional signals, image data had to be transformed. In fish identification applications, otolith contour coordinate points were changed to radial data whereas aging applications moved from gray-scale images to intensity vectors.

The preparation of data may condition the whole analytical process of the recognition system. In fish identification applications, for example, the use of contour radials in a very concave otolith may lead to representations over-passing the contour trace, which would be equivalent to modifying its structures deliberately. In addition, omitting angle information may cause loss of control over homology. In this case, the undesired effect is that the structures being compared do not correspond to analogous positions in other otoliths, developing thus misplaced comparisons.

The proposed normalization framework of otolith contours is only a minimizer of this problem. However, before considering the analysis of full images or the use of very complex invariant transformations one should look at polar or even Cartesian parametrizations, since these may allow us to address this problem while, at the same time, saving computing resources. In such situation, feature extraction mechanisms may need to be modified slightly in order to maintain performance, for example, by using histograms of several dimensions.

7.2.2 Filter Design

In wavelet analysis, multi-resolution approximation is normally seen as a smooth partition of the frequency axis, where the band resolution of the filters increases with decomposition level. In practice, however, this scenario is altered significantly by the process of filter design in order to ensure perfect reconstruction with convolution-sub-sampling operations.

To ensure perfect reconstruction of signals, a low-pass digital filter $\{h[l]\}_{l=0}^{L-1}$ is designed first to construct an interpolating function $\phi(t)$, which is then used to calculate the high-pass filter $\{g[l]\}_{l=0}^{L-1}$) and its corresponding wavelet $\psi(t)$, by means of the two-scale difference equations (3.21) and (3.22). The result of doing the process this way is that the mother wavelet has poor regularity generally speaking, and also fitting its shape to structures commonly found in nature, is difficult. In addition, the indexation of the different dilated and translated versions of this wavelet in the DWPT tree is too complex and for this reason Eq. 3.92 and 3.93 must be used to associate feature parameters.

The choice of particular wavelets is sometimes of importance in certain applications. While complex wavelets are suitable to analyze the time evolution of frequency transients,

because they are capable of separating amplitude and phase components, real wavelets are preferable to detect the sharp time transitions of signal structures, smoothness or even correlation between signals. In this sense, the minimization of the approximation error has been a goal behind the development of some standard basis functions, such as Daubechies wavelets, Coiflets or Symlets, among others (Daubechies 1988); included in most of the software packages. The basics for developing customized real time wavelets are outlined in appendix A.

7.2.3 Translation Dependence

Another issue associated to the normalization of contour data is that contour size, rotation and translation alter coefficient values in the discrete version of the wavelet transforms.

Another alternative to the elliptic Fourier series is known as *spin cycle* (Saito et al. 2002) and consists in minimizing this drawback by including the shifted versions of input signals in the LDB. This is equivalent to the 'bootstrap' validation approach presented in section 4.2.1. In this case, however, the problem may be that the discrimination efficiency of the coefficients may decrease considerably.

Research on this topic is still at an early stage. Despite the computational cost, in our opinion this problem is best addressed by means of time-invariant transforms. In this sense, it may be useful to look for patterns at denser places of the time/scale plane than the dyadic grid,

$$s = a_0^j, \tag{7.1}$$

$$u = k, \tag{7.2}$$

where $1 \leq a_0 \leq 2$. Appendix B summarizes how to compute a discrete version of the continuous wavelet transform (CWT). In particular, The *dyadic wavelet transform* (also referred to as the *algoritme à trous* in French, Holschneider et al. 1989) does not use subsampling and, therefore, maintains translation invariance but a way to track the displacement of coefficients within vectors must be provided.

Despite these problems, wavelets descriptors are considered very flexible tools, suitable for many kinds of signal structures (transients, spikes and other kind of irregularities) and their use can be extended in a wide range of signal processing applications.

7.2.4 Considerations for Multivariate Densities

In general, histograms constitute the simplest form of non-parametric density estimation that is developed nowadays. Its most important parameter corresponds to the bin width h,

since the point of origin t_0 can be determined easily once the lower and upper boundaries are known.

The main problem of multivariate densities is that the optimal bin number grows exponentially. This drawback was coined by Bellman (1961) as the "curse of dimensionality". In our inter-specific/intra-specific *Merluccius* experiments (see chapter 6), the number of bins was relatively too low (oscillating between m=3 and m=8), so there may be reasons to believe that this problem may be behind our unsuccessful classifications.

In the field of density estimation, this problem can be addressed from two perspectives: either we choose a fixed value of V and determine v_m from the data, or 2) we can fix v_m and calculate V. The first group is normally referred to as the *Kernel Density Estimation* in the literature (KDE) (Scott 1985a,b), whereas the second group is known as the *k-Nearest Neighbor* approach (k-NN) (Silverman 1986, Bishop 1995). Both apply to multivariate estimation and their variants address the problem of using small datasets with more, or less, success. They are described in appendix C.

7.2.5 Search Strategies for C Classes

Our current approach may be sufficient for a small number of classes, say from two to four. A major problem of using pairwise measures is that the more signal classes one has in a problem, the more obscure their meaning becomes. In Eq. 5.25 a large value can be due either to few significant terms with negligible majority (a favorable case) or the accumulation of many terms with relatively small values (an unfavorable case). Addressing this problem not only involves the definition of more robust and normalized discrimination measures but may also affect the true number of features necessary to classify data accurately.

Although both bottom-up and top-down *divide-and-conquer* strategies ensure a non-redundant representation time/frequency support for perfect signal reconstruction, there are three main issues which deserve thorough examination in order to develop more robust techniques when working with C classes.

First, is compact support strictly necessary when the goal is to define the best feature models that represent the classes? Or in other words, can redundancy be considered if the accuracy of correct classification improves? In this sense, one could focus the search strategy on any point of the wavelet-packet tree, without considering full nodes.

Secondly, divide-and-conquer methods are too restrictive in relation to feature dependence. That is, they only consider discrimination power in the evaluation of coefficients. However, let us consider the following scenario with five classes and a feature vector of four dimensions, $\mathbf{x} \in \mathbb{R}^4$, showed in pairs of two-dimensional scatter plots (Fig. 7.1):

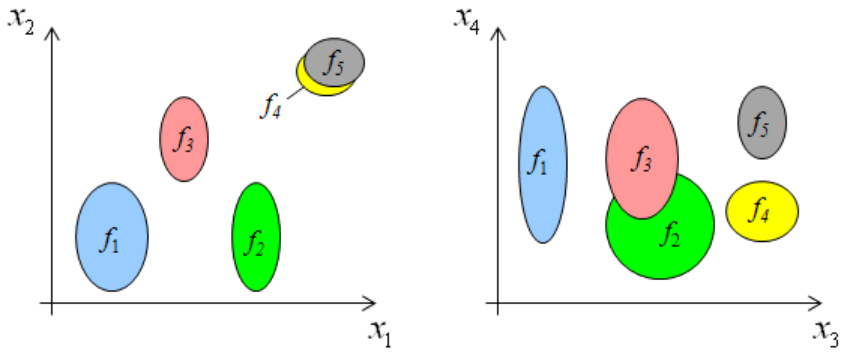


Figure 7.1: Naive sequential feature selection example considering independent features. Only the pair combination $\{x_1, x_4\}$ can obtain full class separation

- The goal is to select the best set of two features identifying the five classes.

- In this situation, any reasonable discriminant measure will rank features according to the sequence: $\delta_{(x1)} > \delta_{(x2)} \approx \delta_{(x3)} > \delta_{(x4)}$. Without doubt, x_1 is the best feature since it divides the class distributions in four groups: f_1, f_2, f_3 and $\{f_4, f_5\}$. x_2 and x_3 have similar efficiency since classes are divided in three groups: $\{f_1, f_2\}$, f_3 and $\{f_4, f_5\}$ for x_2; and f_1, $\{f_2, f_3\}$ and $\{f_4, f_5\}$ for x_3. Finally, x_4 is the worst feature since it can only separate f_4 and f_5.

- The optimal feature set turns out to be $\{x_1, x_4\}$ because x_4 is the only feature that can provide the missing information in x_1: the discrimination between class distributions f_4 and f_5.

- However, if one were to choose two features according to the scores, δ, we would certainly pick $\{x_1, x_2\}$ or $\{x_1, x_3\}$, leaving class distributions f_4 and f_5 non separable.

The strategy fails because it does not account for features of small discrimination value but with useful "complementary information".

In our opinion, *Feature Subset Selection* techniques (FSS) may eventually solve many of these problems. In fact, all that is needed is a robust objective function able to generalize among data and a good search strategy which uses this information as feedback signals to select feature candidates.

In practice, objective functions are divided into two main groups: *filters and wrappers*. Filtering methods respond to the preceding approach (Fig. 7.2a). In this approach, objective functions typically measure inter-class distances, statistical dependence or information-

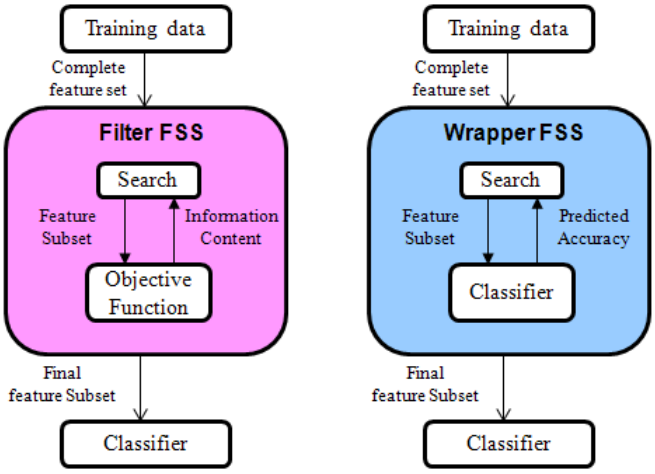


Figure 7.2: Filter-based FSS and Wrapper-based FSS. The objective function of wrapper techniques are based on the predictive accuracy, which is estimated by statistical resampling or cross-validation methods (from Osuna 2005)

theoretic measures; whereas wrappers develop the same functionality by means of built-in classifiers (Fig. 7.2b).

The main advantage of filter approaches lies in their non-iterative computation which, obviously, is much faster than a classifier learning-validation step. The final selected feature set will exhibit more generality providing similar accuracies for a large number of classifiers, but there will be a tendency to select the full feature set, forcing the user to cut-off at an arbitrary number (as in our DLDB). In contrast, recognition rates tend to be higher in wrappers, but only for the classifier under consideration whose interactions are normally tuned for the specific input dataset.

As for the search strategies, there is a large number of methods to be investigated under the *Best-Basis* paradigm. In general, the algorithms can be grouped in three categories: *exponential*, *sequential* and *randomized*. In exponential algorithms, the number of features to evaluate grows exponentially with the dimensionality of the search space (Land and Doig 1960), whereas sequential methods add (or remove) features in each step but tend to become trapped in local minima. This problem is solved partially by randomized algorithms (Banzhaf et al. 1998).

Watanabe and Kaminuma (1988) suggested an interesting framework which can be extended to the Best-Basis paradigm (Saito et al. 2002) in order to address the problem of detecting $C>2$ classes. Instead of constructing a feature extractor for the entire classification problem, consider C sets of two-class problems by reorganizing the partition of the input training set into *c-class* (c) and *non-c-class* ($\bar{c}$). Then, each position of the DWPT tree can

be evaluated according to the previous two-class problem, selecting a total of C coefficients for solving the problem.

7.2.6 Future Proposal

As a concluding remark from this thesis, the following is a proposal to address the preceding issues, and which may be considered in the future:

Algorithm. 7.1: The DLDB2. *Let* $\mathcal{T} = \{\boldsymbol{w}_{j,k,l}\}$ *define the set containing all basis functions corresponding to every position in the wavelet-packet tree, for* $j = 0, \ldots, J$, $k = 0, \ldots, 2^j - 1$ *and* $l = N/2^j - 1$. *Assume that* $\mathcal{B}_c$ *is used to denote the set storing the best features, at each step* c, *initialized to* $\mathcal{B}_0 = \{\phi\}$. *The following procedure describes another possible search strategy:*

- **Step 0:** *Choose a dictionary of orthonormal bases* $\mathfrak{D}$ (*i.e. specify* QMF's *for a wavelet packet dictionary or decide to use either the local cosine dictionary or the local sine dictionary*). *Specify the maximum level of decomposition* J *and a 'bounded' discriminant measure* $\jmath$.

- **Step 1:** *Construct normalized density histograms for both* $\hat{\boldsymbol{f}}_{\boldsymbol{w}}^{(c)}$ *and* $\hat{\boldsymbol{f}}_{\boldsymbol{w}}^{(\bar{c})}$ *distributions, and for each class* $c = 1$, ..., C *compute its associated discrimination value as,*

$$\delta_c(\boldsymbol{w}) = \jmath_{\boldsymbol{w}}(c, \bar{c}), \tag{7.3}$$

  where $\jmath_{\boldsymbol{w}}$ *is the separability measure discriminating among c-class* and *non-c-class* (Eq. 5.24), *associated to basis vector* $\boldsymbol{w}$.

- **Step 2:** *For each class* $c = 1, \ldots, C$ *select the best feature* $\boldsymbol{w}_c$ *isolating class c, according to the following rule*

$$\boldsymbol{w}_c = \arg\max_{\boldsymbol{w} \in \mathcal{T}} (\delta_c(\boldsymbol{w})), \tag{7.4}$$

  and store the selected feature in $\mathcal{B}$ *if needed.* This means $\mathcal{B}_c = \mathcal{B}_{c-1} \cup \boldsymbol{w}_c$, if $\boldsymbol{w}_c \notin \mathcal{B}_{c-1}$ (not stored previously) or set $\mathcal{B}_c = \mathcal{B}_{c-1}$ otherwise.

- **Step 3:** *Use the discriminant basis functions stored in* $\mathcal{B}$ *for constructing classifiers.*

This theoretical approach of LDB is different from the original proposals (Saito 1994, Saito et al. 2002) and the framework proposed in this thesis. It performs *Sequential Forward Selection* (SFS, Land and Doig 1960) over the full wavelet-packet-tree. In some way, it is

meant to address the problem of dimensionality reduction more efficiently. Note that if each class is to be detected all that is needed here is to maximize the selection of one feature per class (one fully discriminating between *c-class* and *non-c-class*). So a maximum vector length of C elements are necessary to solve the problem.

We find that point of view very close to the more philosophical Occam's razor principle (Thorburn, 1915), which suggests that one should proceed first with the simpler theories until simplicity can be traded for greater explanatory power. In this sense, selected feature vectors should not increase in size when one coefficient is capable of handling many classes at a time (note that this possibility is included in the previous formalism since the same $\boldsymbol{w}$ can be repeated for different c). Thus, the final solution should converge to an optimum compression of the data while reaching the maximum possible accuracy of correct classifications.

The explanation of classes in terms of signal structures should also be maintained with this approach. This is possible thanks to the time-frequency properties (Eq. 3.92 and 3.93) of each basis functions, which can be considered independently. Likewise, the generalization properties of discriminant measures are guaranteed, since a two-class normalized measure (i.e. Eq. 5.24) is used.

FSS can be extended within DLDB using other known search methodologies. Table 7.1 summarizes some of the most common methods falling into the three previously mentioned categories and its main performance characteristics.

7.3 Age Estimation

Contrast cancellation and fish growth estimation are two major problems that limit the precision of computer automated fish aging systems nowadays. This fact is evident as observed from the results of the plaice otoliths of this work for manual and automated fish aging systems.

Although AGD seems to be insufficient for demodulating fish growth, it is likely that most of the current limitations are associated to image quality as no information was provided about the light conditions used by TNPC software, making it impossible to eliminate the mutual dependency of both factors in practice.

In this sense such kind of problems are better addressed from the field of computer vision techniques, which allows three-dimensional objects to be characterized by means of images. Specially, the application of *back-projection* methods and *photometric stereo* in static scenes can be crucial to deal with the optical problems present in reflected-light techniques. Another relevant issue arises from the fact the proposed framework of growth demodulation could help in the development of age estimation without using feature extraction and classification tools,

	Accuracy	Complexity	Advantages	Existing methods
Exponential	Always finds the optimal solution	Exponential	High accuracy	Exhaustive Search
				Branch & Bound (B&B)
				Beam Search
Sequential	Good if no backtracking needed	Quadratic $O\left(N^2\right)$	Simple and fast	Forward Sel. (FSF)
				Backward Sel. (SBS)
				Plus-I, minus-R
				Bidirectional Sel. (BDS)
				Forward Floating Sel. (SFFS)
Randomized	Good with proper control parameters	Generally low	Designed to scape local minima	Random Generation plus Sequential Selection
				Simulated annealing
				Genetic Algorithms (GA)

Table 7.1: A brief summary on common search strategies

and thus obtain an estimation of fish age in the domain of real numbers, $\mathbb{R}$.

7.3.1 On Automatic Contrast Cancellation for Aging Technology

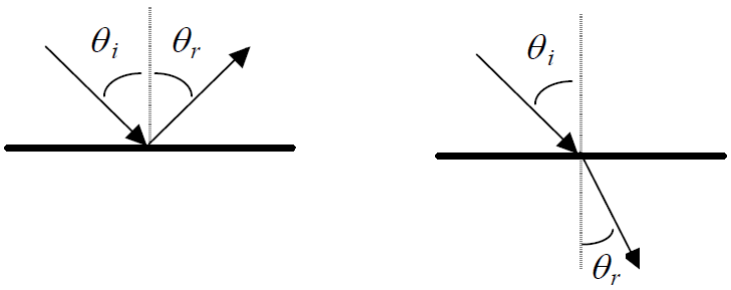
In general, the gray-scale values depend on the illumination conditions where the object surface is recorded, the specifications of the camera used and, finally, the geometric laws which the image acquisition is subjected to.

In physics, the phenomenon of light transmission through objects is normally studied in *geometric optics*, which describes the propagation of light in terms of 'rays' governed by the laws of *reflection* and *refraction*. In general, when the wavelength of the light corresponding to the source is much smaller than the size of the optical system being used, these laws state that the incident light between two transparent materials is divided into reflected and refracted rays (Fig. 7.3). The angle of the reflected ray and incident ray is the same but lies symmetric in relation to the projected plane, whereas the refracted angle is related to the incident angle according to Fermat's principle (Hecht 2000),

$$\frac{\sin\theta_i}{\sin\theta_r} = n_{ab}, \tag{7.5}$$

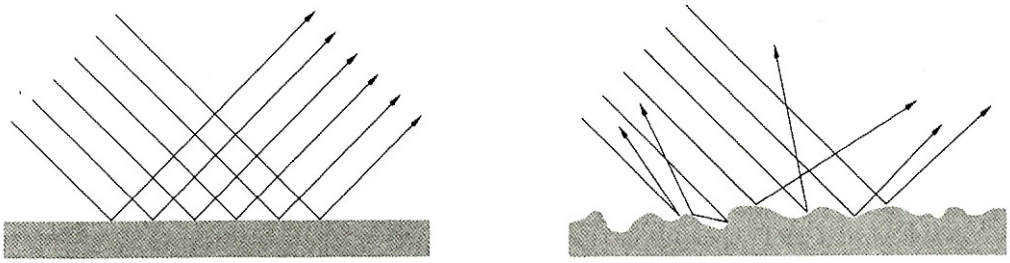
where n_{ab} is known as the refractive index of material a relative to b ($n_{ab} = n_a/n_b$), for a specific color of light.

In turn, reflections of light are divided into *specular* and *diffuse* (Fig. 7.4). Specular reflection is produced from polished surfaces that reflect straight lights in predictive ways. A



(a) (b)

Figure 7.3: (a) Reflected-light. (b) Refracted light



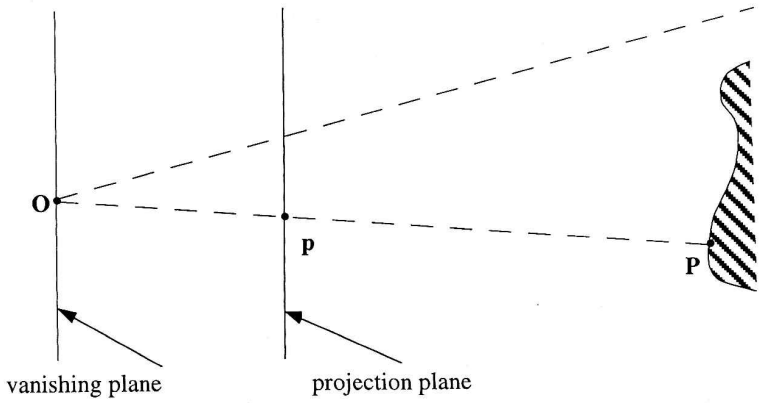
(a) (b)

Figure 7.4: (a) Specular reflection. (b) Diffuse reflection (from Hecht 2000).

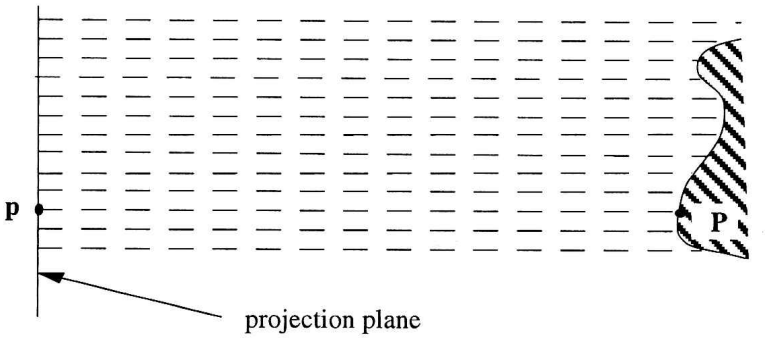
good example is a mirror which is able to reproduce an exact and extrapolated image of an object.

With reflected-light, all surfaces, including the otoliths, will generally exhibit both specular and diffuse reflection. Although white materials can be extremely efficient in giving back the light they receive, reflectance may vary at points of different concavity, which could explain the high contrast variability found in images at the border of the otolith. This scenario needs to be investigated before designing more advanced algorithms. Our main suspicion is that the bias trend observed in the intensity profile is a function of the otolith thickness, so if this thickness can be characterized by means of machine vision techniques, the information can be used to remove the bias of intensity profiles.

In the field of computer vision, surface reconstruction is normally addressed from two main groups of techniques: *back-projection* and *photometric stereo analysis*. Very briefly, the first group is the idealized projection model $E = A(p)$ of points p onto an image E. In this case, there are two possible variants depending on whether the coordinate system is defined as either centered to the camera pinhole (*perspective projection*) or parallel to the image plane (*orthographic projection*) (Fig. 7.5) . Photometric stereo analysis, on the other hand, is based on the philosophy of "shape from shading". The goal is to use two or more images



(a)



(b)

Figure 7.5: Geometric models for surface reconstruction purposes. (a) central projection; (b) orthogonal parallel projection. (from Klette et al. 1998)

of an object with different light conditions in order to determine volume information (Jain et al. 1995, Klette et al. 1998).

In general, otolith-based aging applications have addressed the problem of image quality from the perspective of graphical enhancement tools, perhaps based on the believe that light conditions are easily controlled for static images. However, in our opinion, it is reasonable to expect that the characterization of age structures from entire otoliths may need the consideration of thickness and surface volume information in order to determine the amount of light that is projected from outside the camera focus, specially at the tail of the intensity profiles which may be the main cause for high contrast variability and thus the steep slope at the right end of the intensity profile.

7.3.2 About the Use of Classifiers for Aging Purposes

If we assume that the problem of estimating the true demodulation parameters β^* can be solved, then a simple, low-cost and, if compared to current aging systems, less time-consuming technique exists to determine fish age. As this computation does not require feature extraction or classification tools, the full aging system will be filter-based.

Let us recall that $\hat{f}_\beta(u)$ was defined in Eq. 5.27 as a density estimate for the periodicity of age structures, constructed from the lags $u_{m,\beta}$. Let $\boldsymbol{f}_i^*$ denote the optimal density of intensity profile $\boldsymbol{i}$ calculated from its optimal demodulation parameters β_i^*. Our following hypothesis establishes the rule for inferring the age:

Hypothesis 3. Definition of Fish Age. *Given the optimal demodulation parameters β_i^* and the density $\boldsymbol{f}_i^*$ corresponding to intensity profile $i[n]$ for $(N \in \mathbb{Z})$, then, the position defined by*

$$u_i^* = \frac{1}{\|\boldsymbol{f}_i^*\|^2} \sum_j k_j \left| f_i^*[u] \right|^2 \text{ for } u \in \mathbb{R}, \tag{7.6}$$

will correspond to the length of the most possible regular year period along the intensity profile and the rule

$$r(\boldsymbol{i}) = N/u_i^*, \tag{7.7}$$

infers the estimated fish age, where $N/u_i^* \in \mathbb{R}$.

In other words, the exact fish age is simply the relation between the total number of samples of its intensity profile and the length of one year period, given by the centroid of $f_i^*[u]$. This length, expressed on a number-of-sample basis, is associated to the demodulation that best relocates all age structures regularly along the intensity profile.

This new definition of fish age has serious implications compared to current methodology, as it quotes that if the regular periodicity can be estimated from the growth function, the exact age is simply the division of two values. Therefore, the only issue for developing methods that automatically estimate fish age is concerned with the problem of designing a robust and automatic demodulation method that adapts to the true fish growth.

7.4 Final Remark

Ensuring consistency between discriminatory measures and classification accuracy when looking for relevant features in the domain is crucial in order to provide generalization in pattern

recognition applications. This performance was demonstrated in two-class identification experiments, where the inter-specific and intra-specific studies of *merluccius* spp allowed the otolith structures to be interpreted in terms of structural components and their ecological differences.

The developed tools complement those of classical methodology. Although results demonstrate the contributions towards the automated detection of signal and image patterns, the methods are not intended to be contenders of the classical ones. The bridge between the structural and statistical methods was decisive for evaluating the relevance of selected structural components. In this sense, we believe that combining both groups of tools is essential to detect signal patterns and to develop more robust feature extraction systems in the future.

Class distributions have demonstrated themselves to be primary information source for achieving measure normalization. This behavior is crucial in the development of more generalized feature extractors. Here, the main goal is to derive a system capable of obtaining the shortest possible feature vector that explains the class differences (Occam's razor principle).

In this sense, our contention is that the functionality of future extractor systems must bridge the gap between linear and nonlinear classifiers. This means that the method must be automatic adaptable from an evaluation based on discriminant measures, to an evaluation based on decision boundaries when class overlapping is detected. In addition, levels of priority between both performances must be established to ensure that the shortest possible feature length is able to detect all the classes.

The suggestions outlined in this chapter should allow new automatic pattern recognition tools to be developed towards this goal. They address multi-class detection, time translation and the description of time-frequency atoms, among other issues.

The estimation of fish age, on the other hand, may not require the use of complex classifiers or feature extraction systems at all, as long as growth demodulation is addressed correctly for each individual otolith, but image preprocessing will have to be previously addressed from the more technical (or even mechanical) issues of acquisition systems and machine vision applications in order to reduce the current problems of contrast variation.

Part V

Appendices

Appendix A

Wavelet Design

A.1 Introduction

From a signal processing point of view, a wavelet is a bandpass filter. In the dyadic case given in Eq. 3.17 it is actually an octave band filter. Therefore, the wavelet transform can be interpreted as a constant Q-filtering with a set of octave filters, followed by sampling at the respective Nyquist frequencies, corresponding to the bandwidth of the particular octave band.

The connection between wavelets, filter-banks and multiresolution signal processing (Mallat 1990, Meyer 1990) is shown in this appendix. Most of the theory of this framework is described in the excellent work of Vetterli (1992) and is reproduced here since it constitutes a main reference for the construction of customized real time wavelets and filter-bank design. Some methods based on algebraic structures will be reviewed in the last section. These include, paraunitary filter-banks and, regular filter design and Diophantine equations (Vaidyanathan 1990, Vetterli 1992).

A.2 Multiresolution Signal Processing

Let V_0 define the space of all band-limited signals with frequencies in the interval $(-\pi, \pi)$. Similarly, call V_{-1} the space of band-limited signals in $(-2\pi, 2\pi)$. Then, the set of functions

$$\phi(t-k) = \text{sinc}(t-k) = \frac{\sin(\pi(t-k))}{\pi(t-k)},\ k \in \mathbb{Z} \tag{A.1}$$

forms an orthonormal basis for V_0. Clearly, the set $\sqrt{2}\text{sinc}(2t-k)$, $k \in \mathbb{Z}$ is an orthonormal basis for V_{-1}. In particular, if $x(t) \in V_0$, then $x(2t) \in V_{-1}$. Now, call W_0 the orthogonal space

of bandpass signals with frequencies in the interval $(-2\pi, -\pi) \cup (\pi, 2\pi)$, which complements V_0 in V_{-1},

$$V_{-1} = V_0 \cup W_0. \tag{A.2}$$

In other words, V_{-1} is equivalent to V_0 with some added details corresponding to W_0. For completeness $\cos(\pi t)$ has to be included in V_0, and $\sin(\pi t)$ in W_0. From the above it is clear, by scaling, that if V_i represents the frequency interval $(-2^{-j}\pi, 2^{-j}\pi)$, then

$$V_j \subset V_{j-1}\ i \in \mathbb{Z} \tag{A.3}$$

$$V_{j-1} = V_j \cup W_j\ i \in \mathbb{Z} \tag{A.4}$$

where W_i represents the interval $(-2^{-j+1}\pi, -2^{-j}\pi) \cup (2^{-j}\pi, 2^{-j+1}\pi)$. Moreover, we have

$$\cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \cdots$$

and by iterating A.4 it holds that the union

$$V_j = W_{j+1} \cup W_{j+2} \cup W_{j+3} \cup \cdots \tag{A.5}$$

for all W_j's, $j = 1, \ldots, \infty$ is equivalent to the space of square integrable functions, $\mathrm{L}^2(\mathbb{R})$, band limited to $(-2^{-j+1}\pi, 0) \cup (0, 2^{-j+1}\pi)$.

In the sampled version of V_{-1}, $\phi(t)$ can be written as the interpolation, by $\phi(2t)$, of the perfect half-band low-pass filter with impulse response;

$$c[n] = 2^{-1/2}\frac{\sin(\pi n/2)}{\pi n/2} \equiv \text{discrete halfband filter}\ . \tag{A.6}$$

That is,

$$\phi(t) = \sum_{n=-\infty}^{\infty} c[n]\,\phi(2t-n)\ \ -\infty < n < \infty \tag{A.7}$$

for $n \in \mathbb{Z}$. In other words, $\phi(t)$ derives an approximation of V_0 in V_{-1}. The orthogonal complement, W_0, will then be given by the half-band high-pass signals. In the sampled domain, this is equivalent to modulating the low-pass filter (Eq. A.6) by $(-1)^n$, and shifting the vector sequence by one (in order to include $\sin(\pi t)$). Thus, $\psi(t)$ is the interpolation thereof, that is,

$$\psi(t) = \sum_{n=-\infty}^{\infty} (-1)^n c[-n+1]\,\phi(2t-n). \tag{A.8}$$

It holds that $\phi(t-k) \perp \psi(t-k)$ since they cover disjoint regions of the spectrum and, therefore, $\langle \phi(t-k), \psi(t-l) \rangle = 0$. In addition, $\langle \phi(t-k), \phi(t-l) \rangle = \delta_{kl}$ because the translates $\phi(2t-n)$ are even and the sign change is canceled. Likewise, $\langle \psi(t-k), \psi(t-l) \rangle = \delta_{kl}$, so $\psi(t)$ (and its integer translates) form an orthonormal basis for W_0. As such, in this example the wavelet for the "sinc" scaling function is given by $\psi(t)$.

Equations (A.7) and (A.8) establish the basic framework for wavelet filter design. In particular, if we manage to find an interpolating function, $\phi(t)$, so that its integer translates form an orthonormal basis in V_0, then we can find the coefficients $c[n]$ such that $V_0 \subset V_{-1}$. Then, $\psi(t)$ and its integer translates will form an orthonormal basis for W_0. So, the set $\left\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j x - k); j \in \mathbb{Z}\right\}$ will constitute an orthonormal basis for $\mathrm{L}^2(\mathbb{R})$, following expression A.5.

A.2.1 Discrete Signals

Assume now that we deal with discrete time sequences $x[n]$, $n \in \mathbb{Z}$ which are square summable (they belong to the space$l^2(\mathbb{Z})$). One way to derive a coarse half-resolution approximation to the original sequence, $\overline{x}[n] \equiv x(2t)$, is by filtering with a half-band lowpass filter, followed by sub-sampling by 2 (discarding every odd-indexed sample). Assuming for the sake of simplicity that the filter is FIR, convolving the filter impulse response $h_0[n] = (h_0[0], h_0[1], \ldots, h_0[L-1])$ and sub-sampling by 2, corresponds to the matrix multiplication of the infinite-length signal vector $(\ldots, x[-1], x[0], x(1), \ldots)$ by

$$\boldsymbol{H_0} = \begin{pmatrix} & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ \ddots & h_0[L-2] & h_0[L-1] & \cdots & \cdots & h_0[0] & 0 & 0 & \ddots \\ & 0 & 0 & h_0[L-1] & \cdots & h_0[2] & h_0[1] & h_0[0] & \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}. \tag{A.9}$$

Let us assume that the impulse response and its shifted versions (the rows of the above matrix) form an orthonormal set, that is,

$$\langle h_0[n-2l], h_0[n-2k] \rangle = \delta_{kl}, \tag{A.10}$$

for $k, l \in \mathbb{Z}$. In matrix notation the equivalent of (A.30) is the relation $\boldsymbol{H_0 H_0^*} = \boldsymbol{I}$. The projection of the original sequence $x[n]$ onto the subspace spanned by the rows of $\boldsymbol{H_0}$ is given

by $\boldsymbol{H_0^* H_0 x}$, where multiplication by $\boldsymbol{H_0^*}$ corresponds to up-sampling by 2 followed by convolution with a filter of impulse response $\widetilde{h}_0[n] = (h_0[L-1], h_0[L-2], \ldots, h_0[1], h_0[0])$, (the time reversed input response of $h_0[n]$). Note that in order for the set $\{h_0(n-2k), k \in \mathbb{Z}\}$ to form an orthonormal basis L has to be even, or else $\langle h_0[n], h_0[n-L+1]\rangle \neq 0$ unless either $h_0[0]$ or $h_0[L-n-1]$ was zero.

Note that if V_0 is the subspace spanned by the rows of $\boldsymbol{H_0}$, the filter with impulse response $h_1[n] = (-1)^{1+n} h_0[L-1-n]$ and its shifted versions will form an orthonormal basis for W_0. The orthogonality of $\boldsymbol{h_0}$ and $\boldsymbol{h_1}$ with respect to even shifts is easily verified, because of the sign in $\boldsymbol{h_0}$:

$$\langle h_1[n-2l], h_0[n-2k]\rangle = 0. \tag{A.11}$$

for $k, l \in \mathbb{Z}$. In matrix notation, this is equivalent to $\boldsymbol{H_0 H_1^*} = 0$, where now $\boldsymbol{H_1^*}$ is the matrix based on the reverse impulse response of $h_1[n]$. Then, from the orthonormality condition, $V_0 \perp W_0$ it holds that

$$\langle h_1[n-2l], h_1[n-2k]\rangle = \delta_{kl}, \tag{A.12}$$

for $k, l \in \mathbb{Z}$; since the sign of $h_1[n]$ is canceled.

In other words, based on an orthonormal basis for V_0, we constructed an orthonormal basis for W_0, just as in the continuous time case. Again in matrix notation, $\boldsymbol{H_1 H_1^* = I}$. The coarse half-resolution approximation of the original signal is given by

$$\overline{\boldsymbol{x}} = \boldsymbol{H_0^* H_0 x}, \tag{A.13}$$

whereas the difference (or "detail") signal is determined as

$$\boldsymbol{d} = \left(\boldsymbol{I - H_0^* H_0}\right) = \boldsymbol{H_1^* H_1 x}. \tag{A.14}$$

But, because it is a perfect reconstruction system ($\boldsymbol{x} = \overline{\boldsymbol{x}} + \boldsymbol{d}$) it follows that $\boldsymbol{H_0^* H_0} + \boldsymbol{H_1^* H_1} = \boldsymbol{I}$ and, therefore,

$$\boldsymbol{H_1^* H_1} = \boldsymbol{I} - \boldsymbol{H_0^* H_0}. \tag{A.15}$$

Fig. A.1 recapitulates the above relationships in standard digital processing notation, using filters and sampling rate changes to denote operators so far. As in the continuous case, the direct sum of all W_i's is the space of square integrable functions $l^2(\mathbb{Z})$. The decomposition of V_{-1} into W_0, W_1, W_2, etc, is essentially a wavelet transform on discrete sequences (DWT). Actually, if the filter $h_0[n]$ is an ideal half-band low-pass filter given by

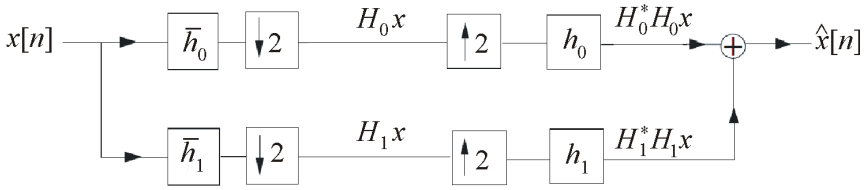


Figure A.1: Decomposition of V_{-1} into V_0 using multirate filters, and recombination to achieve perfect reconstruction. $\boldsymbol{H}_0^*\boldsymbol{H}_0\boldsymbol{x}$ is the projection of the signal onto V_0 and $\boldsymbol{H}_1^*\boldsymbol{H}_1\boldsymbol{x}$ is the projection onto W_0.

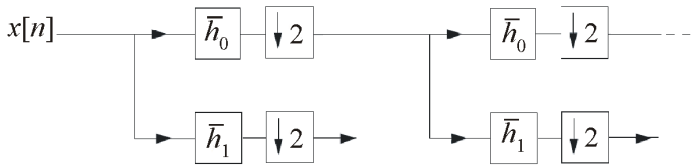


Figure A.2: Discrete wavelet transform on sequences. The half-band low-pass and high-pass filters are $h_0[n]$ and $h_1[n]$, respectively, and $2\downarrow$ means sub-sampling by 2 (dropping odd-indexed samples).

Eq. A.6, then $h_1[n]$ is the ideal half-band high-pass filter. Thus, if V_{-1} is the space of signals band limited to $(-2\pi, 2\pi)$, then V_0 and W_0 are the spaces of signals band limited to $(-\pi, \pi)$ and $(-2\pi, -\pi) \cup (\pi, 2\pi)$, respectively. Thus, by iteration, the discrete system in Fig. A.2 computes the DWT exactly into octave bands. However, in contrast to the continuous case there is a "maximum" resolution in the discrete case given by the original sampling rate at V_{-1}.

A.2.2 FIR Filter Banks and Compactly Supported Wavelets

For the function $\phi(t) = \sin(\pi t)/\pi t$, equation (A.6) establishes the coefficient values of the filter that are needed to compute the wavelet transform. Thus, the sequence of the low-pass filter are samples from the scaling function, $h_0[n] = 1/\sqrt{2}\phi(n/2)$ and, subsequently, $h_1[n] = (-1)^n h_0[L-1-n]$. However, this choice is unpractical because it involves the use of an ideal infinite impulse response low-pass filter (IIR) of slow decay as $t \to \infty$.

The orthogonality conditions of compact support,

$$\langle \phi(t-l), \phi(t-k) \rangle = \delta_{kl} \tag{A.16}$$

$$\langle \psi(t-l), \psi(t-k) \rangle = \delta_{kl} \tag{A.17}$$

$$\langle \phi(t-l), \psi(t-k) \rangle = 0 \tag{A.18}$$

can also be accomplished by means of FIR filters, provided that the linear combination of the set $\{\phi(2t-n)\}$ converges to the continuous $\phi(t)$ for some finite n. Let $\phi(t)$ and $\psi(t)$ form an orthonormal basis which obey two-scale difference equations (A.7) and (A.8). From (A.16) we obtain

$$\langle \phi(2t-k), \phi(2t-l) \rangle = \frac{1}{2}\delta_{kl}. \tag{A.19}$$

Now, using (A.7) and (A.19), (A.16) can be written as (with $n' = n + 2l = n + 2k = m'$)

$$\begin{aligned}
&\langle \phi(t-l), \phi(t-k) \rangle \\
&\quad = \left\langle \sum_n c[n]\,\phi(2t-2l-n), \sum_n c[n]\,\phi(2t-2k-n) \right\rangle \\
&\quad = \left\langle \sum_{n'} c[n'-2l]\,\phi(2t-n'), \sum_{m'} c[m'-2k]\,\phi(2t-m') \right\rangle \\
&\quad = \sum_n c^2[n]\,\langle \phi(2t-n'), \phi(2t-m') \rangle \\
&\quad = \frac{1}{2}\sum_n c^2[n]\,\delta_{kl}
\end{aligned} \tag{A.20}$$

from which it follows that $\|c[n]\| = \sqrt{2}$.

In other words, the discrete filter of impulse response $h_0[n] = c[n]/\sqrt{2}$ is orthogonal to its even translates, and with $h_1[n] = (-1)^n h_0[L-1-n]$, an orthogonal perfect reconstruction FIR filter bank is obtained. Thus, compactly supported wavelets lead to perfect reconstruction FIR filter banks. The converse, however, is only true under certain constraints which we now review, since it constitutes the main basis for the implementation and design of compactly supported wavelets.

Considering the DWT in Fig. A.2, it can be observed that the lower branch corresponds to an infinite cascade of filters $h_0[n]$ followed by sub-sampling by 2. In the z-domain, sub-sampling a signal of z-transform $X(z)$ by 2 results in a new signal with z-transform,

$$X_{down}(z) = \frac{1}{2}\left[X\left(z^{1/2}\right) + X\left(-z^{1/2}\right)\right]. \tag{A.21}$$

Furthermore, sub-sampling by 2 followed by filtering with $H(z)$ is equivalent to filtering with $H(z^2)$ (the up-sampled filter) followed by sub-sampling by 2, so the cascade of j filter-downsampling operations is also equivalent to a filter $H^{(j)}(z)$ with z-transform,

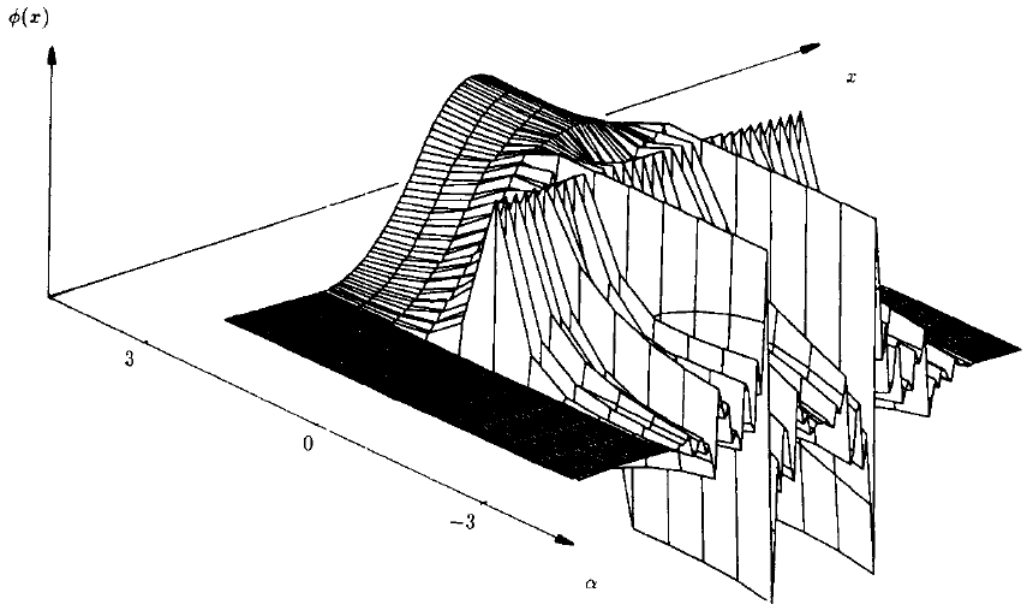


Figure A.3: Scaling function generated by using (A.23) for $h_0 = [1, \alpha, \alpha, 1]$ and $\alpha \in \{-3, 3\}$.

$$H^{(j)}(z) = \prod_{l=0}^{j-1} H\left(z^{2^l}\right) \quad j = 1, 2 \cdots, \tag{A.22}$$

followed by sub-sampling by 2^j. Defining $H^{(0)}(z) = 1$ and assuming that $H(z)$ has length L, it holds from A.56 that the length of $H^{(j)}(z)$ is $L^{(j)} = (2^j - 1)(L - 1)$. Of course, as $j \to \infty$, we obtain $L^{(j)} \to \infty$. Let us consider the function $f^{(j)}(t)$, instead of the discrete time filter, which is constant on intervals of length $1/2^j$. That is,

$$f^{(j)}(t) = 2^{j/2} h^{(j)}[n] \qquad \text{for } n/2^j \leq t \leq (n+1)/2^j. \tag{A.23}$$

Clearly, $f^{(j)}(t)$ is supported in the interval [0,L-1]. The normalization factor $2^{j/2}$ ensures that if $\sum \left(h^{(j)}[n]\right)^2 = 1$ then $\int \left(f^{(j)}(t)\right)^2 dt = 1$ as well. It can be shown that $\left\|h^{(j)}\right\|_2 = 1$ when $\left\|h^{(j-1)}\right\|_2 = 1$.

An important issue here is to find out what the function $f^{(j)}(t)$ converges to as $j \to \infty$. Habitually, regular continuous functions are often preferred when constructing wavelets of compact support, perhaps including derivatives also. This can be achieved imposing certain regularity constraints in the initial filter $H^{(1)}(z)$. However, one should be careful since, for example, changing coefficient signs may lead to shapes of quite different irregularities (see Fig. A.3).

First, assume that the filter $H(z)$ has a zero at $\omega = \pi$, or $H\left(e^{i\omega}\right) = 0$. This is quite reasonable if $H(z)$ is to be a half-band low-pass filter; in fact such a zero is necessary if $f^{(j)}(t)$ is to converge to a continuous function when $j \to \infty$ (Rioul 1993). This together with the fact that impulse responses are orthogonal to even translates is equivalent to $\sum h[n] = H(1) = \sqrt{2}$. Define the polynomial $m_0(z) = \left(1/\sqrt{2}\right) H(z)$, where $m_0(1) = 1$, and factorize into its roots at π and the remainder polynomial $R(z)$ as,

$$m_0(z) = \left[\left(1 + z^{-1}\right)/2\right]^N R(z). \tag{A.24}$$

where N denotes number of vanishing moments. Now, call B the supremum of $|R(z)|$ on the unit circle,

$$B = \sup_{\omega \in [0,2\pi]} \left|R\left(e^{i\omega}\right)\right|. \tag{A.25}$$

Then, the following result from Daubechies (1988) holds:

Remark. **A.1:** *If $B < 2^{N-1}$, then the piecewise function, $f^{(j)}(t)$, defined in* (A.23) *converge point-wise to a continuous function, $f^{(\infty)}(t)$.*

This regularity condition is sufficient to ensure point-wise convergence by means of digital FIR filters and is often used as a test in the construction of wavelets. However, a more accurate verification is based on the estimation of a regularity index β such that $\psi(t)$ and $\phi(t) \in C^\beta$ (Daubechies 1988, Rioul 1993). Such methods are potentially useful, in particular, if a filter fails to test condition A.1. The main reason is that a design criterion can be made in order to place a maximum number of zeros at π so that an efficient signal compression can be reached through the wavelet transform (note that if $m_0(z)$ has a zero of order N at z=-1 then the wavelet will contain N consecutive vanishing moments, that is, $\int t^k \psi(t)\, dt = 0,\ k = 0, 1, \ldots N-1$).

A.2.2.1 Bases of Orthonormal Wavelets Constructed from Filter Banks

Using (A.22) and (A.23) we get

$$f^{(j)}(t) = 2^{j/2} \sum_{m=0}^{L-1} h_0[m]\, h_0^{(j-1)}\left[n - 2^{j-1}m\right], \qquad \text{for } n/2^j \le t < (n+1)/2^j.$$

The term $h_0^{(j-1)}\left[n - 2^{j-1}m\right]$ can be written as a function of $f^{(j-1)}(2t - m)$ by means of (A.23) using the change of variable $n' = n + 2^{j-1}l$. That is,

$$f^{(j-1)}(2t - m) = 2^{(j-1)/2} h_0^{(j-1)}\left[n' - 2^{j-1}m\right], \tag{A.26}$$

for $n'/2^j \leq t < (n'+1)/2^j$ and, therefore,

$$f^{(j)}(t) = 2^{1/2} \sum_{m=0}^{L-1} h_0[m] f^{(j-1)}(2t-m). \tag{A.27}$$

Since $f^{(j)}(t)$ tends to a continuous limit function $\phi(t)$ as $j \to \infty$, the preceding relation satisfies a two-scale difference equation,

$$\frac{1}{\sqrt{2}}\phi(t/2) = \sum_{n=0}^{L-1} h_0[n]\phi(t-n) = \boldsymbol{h}_0 * \boldsymbol{\phi}. \tag{A.28}$$

Similarly, the bandpass function are obtained as

$$\frac{1}{\sqrt{2}}\psi(t/2) = \sum_{n=0}^{L-1} h_1[n]\phi(t-n) = \boldsymbol{h}_1 * \boldsymbol{\phi}. \tag{A.29}$$

These fundamental expressions establish the relations between the basis functions of two consecutive octave scales. Furthermore, if $h_0[n]$ and $h_1[n]$ are orthogonal with respect to even shifts,

$$\langle h_i[n-2l], h_j[n-2k] \rangle = \delta_{ij}\delta_{kl}, \tag{A.30}$$

for $k, l \in \mathbb{Z}$, it can be shown that $\phi(t-k)$ is orthogonal to integer translates of itself, $\forall k$, an $\psi(t)$ is orthogonal across scales.

Let us assume that $f^{(j)}(t)$is orthogonal at the jth level,

$$\left\langle f^{(j)}(t-l), f^{(j)}(t-k) \right\rangle = \delta_{kl}. \tag{A.31}$$

Then, it holds that $f^{(j+1)}(t)$ is also orthogonal,

$$\begin{aligned}
&\left\langle f^{(j+1)}(t-l), f^{(j+1)}(t-k) \right\rangle \\
&\qquad = 2\sum_n \sum_m h_0[n] h_0[m] \left\langle f^{(j)}(2t-2l-n), \right. \\
&\qquad\quad \left. f^{(j)}(2t-2k-m) \right\rangle \\
&\qquad = \sum_n h_0[n] h_0[n+2l-2k] \\
&\qquad = \delta_{kl}.
\end{aligned}$$

Hence, by induction equation (A.31) holds for all j. So in the limit ($j \to \infty$),

$$\langle \phi(t-l), \phi(t-k) \rangle = \delta_{kl}. \tag{A.32}$$

The other orthogonality relations $\psi(t)$ and $\phi(t)$ follow easily from (A.30),

$$\langle \psi(x-l), \psi(x-k) \rangle = \delta_{kl} \tag{A.33}$$

$$\langle \phi(x-l), \psi(x-k) \rangle = 0 \tag{A.34}$$

Finally, across one scale it holds that

$$\begin{aligned} \langle \psi(t-l), \psi(2t-k) \rangle &= \left\langle \sum_n h_1[n] \phi(2t-2l+n), \psi(2t-k) \right\rangle \\ &= \sum_n h_1[n] \langle \phi(2t-2l+n), \psi(2t-k) \rangle \\ &= 0. \end{aligned}$$

Similar reasoning shows that across several scales $\psi(2^j t - l)$ is orthogonal to $\psi(2^i t - k)$ for $i \neq j$ and $k \neq l$. Additionally, completeness is required in order that $\{2^{-j}\psi(2^j t - l); j, l \in \mathbb{Z}\}$be an orthonormal basis for $L^2\mathbb{R}$. This is proved in (Daubechies 1988), so any signal $x(t) \in L^2\mathbb{R}$ can be represented as compositions from the basis set $\left\{2^{-j/2}\psi(2^j t - l), j, l \ni \mathbb{Z}\right\}$:

$$x(t) = \sum_j \sum_k \left\langle 2^{-j/2}\psi\left(2^j t - k\right), x(t) \right\rangle 2^{-j/2}\psi\left(2^j t - k\right).$$

In summary, if the digital filters $h_0[n]$ and $h_1[n]$ and their even translates form an orthonormal set in $l^2\mathbb{Z}$, orthogonal wavelet bases in $L^2\mathbb{R}$ of different integer translates can be generated. The resulting filter bank can be iterated infinitely without loosing this orthogonality condition, making it possible to generate wavelet bases for tree-structured transforms with perfect reconstruction capabilities.

A.2.3 General FIR Perfect Reconstruction Filter Banks

Assume that we have a filter bank as in Fig. A.1 with analysis filters $H_0(z)$ and $H_1(z)$, and general synthesis filters $G_0(z)$ and $G_1(z)$ (instead of the orthogonal synthesis filters $H_0^*(z)$ and $H_1^*(z)$, respectively). Knowing that up-sampling by 2 corresponds to simply replacing z by z^2 in the z-transform, it is easily seen (by using A.21) that the output of the analysis/synthesis system is given by

$$\hat{X}(z) = \frac{1}{2}\begin{bmatrix} G_0(z) & G_1(z) \end{bmatrix}\begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix}\begin{bmatrix} X(z) \\ X(-z) \end{bmatrix}. \tag{A.35}$$

Let us denote the above 2×2 matrix $\boldsymbol{H_m}(z)$, where m indicates that it contains modulated versions of the filters $H_0(z)$ and $H_1(z)$. To perfectly reconstruct signals we need to cancel the contribution of the alias term, represented by $X(\text{-}z)$. This can be done by relating the synthesis filters to the analysis filters as follows,

$$\begin{bmatrix} G_0(z), & G_1(z) \end{bmatrix} = C(z)\begin{bmatrix} H_1(-z) & H_0(-z) \end{bmatrix}. \tag{A.36}$$

Note that

$$\begin{aligned} \det[\boldsymbol{H_m}(z)] &= H_0(z)H_1(-z) - H_0(-z)H_1(z) && \text{(A.37)} \\ &= P(z) - P(-z) && \text{(A.38)} \end{aligned}$$

where $P(z) = H_0(z)H_1(-z)$. At this point we introduce the so called poly-phase filter notation,

$$H_i(z) = H_{i0}\left(z^2\right) + z^{-1}H_{i1}\left(z^2\right), \quad (i = 0, 1), \tag{A.39}$$

where $H_{i0}(z)$ contains the even-indexed coefficients of the filter $H_i(z)$, and $H_{i1}(z)$ the odd ones. Thus, using (A.21) it follows that

$$\begin{bmatrix} H_{00}(z^2) & H_{01}(z^2) \\ H_{10}(z^2) & H_{11}(z^2) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & z \end{bmatrix},$$

or equivalently,

$$\boldsymbol{H_p}\left(z^2\right) = 2^{-1}\boldsymbol{H}_m(z)\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & z \end{bmatrix},$$

where $\boldsymbol{H_p}(z)$ is called the poly-phase matrix. In particular,

$$\det[\boldsymbol{H_m}(z)] = -2z^{-1}\det\left[\boldsymbol{H_p}\left(z^2\right)\right] \tag{A.40}$$

and

$$\begin{aligned} \det[\boldsymbol{H_p}(z)] &= H_{00}(z)H_{11}(z) - H_{01}(z)H_{10}(z) \\ &= \frac{1}{2}z^{1/2}\left[P\left(z^{1/2}\right) - P\left(-z^{1/2}\right)\right]. && \text{(A.41)} \end{aligned}$$

It is a well known fact that to obtain perfect reconstruction with FIR synthesis filters, after a FIR analysis, the following condition (Smith and Barnwell 1984, Vetterli 1986, Vaidyanathan and Hoang 1988) is necessary,

$$\det\left[\boldsymbol{H_m}\left(z\right)\right] = cz^{-2l-1} \quad l \in \mathbb{Z}. \tag{A.42}$$

Note that $\det\left[\boldsymbol{H_p}\left(z\right)\right]$ is thus a pure delay as well. In order to make results less arbitrary, the constant is fixed to c=2. From this and (A.37) it follows that $P(z)$ can have only a single nonzero odd-indexed coefficient:

$$P\left(z\right) - P\left(-z\right) = 2z^{-2l-1}. \tag{A.43}$$

A polynomial $P(z)$ satisfying this constraint is termed a *valid* polynomial. A factorization such that $P\left(z\right) = H_0\left(z\right)H_1\left(-z\right)$ will give a possible FIR perfect reconstruction filter bank, and $H_0\left(z\right)$ and the complementary filter $H_1\left(z\right)$ will form a perfect reconstruction set. Additionally, note that $H_0\left(z\right)$ and $H_1\left(-z\right)$ are interchangeable. This constraint establishes two possible scenarios of filter design: 1) the *factorization method* where the goal consists in finding a valid $P(z)$ satisfying some design criteria, and then factoring it into $H_0\left(z\right)$ and $H_1\left(-z\right)$; or 2) the *complementary filter method*, which starts with a filter $H_0\left(z\right)$ and then solves a system of linear equations to find the filter $H_1\left(z\right)$ leading to a valid $P(z)$. In either case, once $P(z)$ is found and factored in terms of the analysis filters $H_0\left(z\right)$ and $H_1\left(-z\right)$ the synthesis filters follow directly from (A.36), with $C\left(z\right) = cz^{-l}$.

A.2.3.1 Orthogonal or Paraunitary Filter Banks

In the subsections above, we have shown that orthogonal filters must be orthonormal with respect to even translates. Using filtering notation, this means that the even terms of the autocorrelation $H_i\left(z\right)H_i\left(z-1\right)$ are all zero, with the exception of the central one which equals unity (for normalized filters). Thus, the condition (A.30) becomes

$$H_i\left(z\right)H_i\left(z^{-1}\right) + H_i\left(-z\right)H_i\left(-z^{-1}\right) = 2 \quad i \in \left\{0,1\right\}. \tag{A.44}$$

Furthermore, the two filters $H_0\left(z\right)$ and $H_1\left(z\right)$ are orthogonal to each other at even translates (A.11), so the even terms of the cross correlation are all zero:

$$H_0\left(z\right)H_1\left(z^{-1}\right) + H_0\left(-z\right)H_1\left(-z^{-1}\right) = 0. \tag{A.45}$$

Considering that $h_0\left[n\right]$ is of even length, the term $H_0\left(z\right)H_1\left(z^{-1}\right)$ only presents odd coefficients, so it can be written as

$$H_0(z) H_1\left(z^{-1}\right) = z^{-2n-1} Q\left(z^2\right),$$

which means that zeros must appear in pairs at $z = \alpha$ and $z = -\alpha$, so that the low-pass and high-pass filters form an orthogonal basis in $l^2(\mathbb{Z})$. However, $H_0(z)$ alone cannot have such a pair of zeros if perfect reconstruction is to be possible (see remark B.2.2 below), so for every zero at $z = \alpha$ in $H_0(z)$ there must be a corresponding zero at $z = -\alpha$ in $H_1(z^{-1})$, or equivalently, $H_1(z)$ must have a zero at $z = -1/\alpha$, which lead to

$$H_1(z) = z^{-2k-1} H_0\left(-z^{-1}\right). \tag{A.46}$$

In other words, the orthogonal high-pass filter is the alternating flip of the low-pass ones, varying the sign of the odd-indexed elements: $h_1[n] = (-1)^{n+1} h_0[L-1-n]$. Additionally, the low-pass and high-pass synthesis FIR filters are given by $C(z) = cz^{-l}$ so they correspond to the reversed versions of the analytic ones: $g_0[n] = \overline{h}_0[n]$ and $g_1[n] = \overline{h}_1[n]$, respectively.

This constraint establishes the relation between the low-pass and high-pass filters so that their impulse responses and even translates form an orthogonal basis for $l^2(\mathbb{Z})$. It also means that $H_1(z)$is related to the poly-phase matrix as

$$H_1(z) = -z^{-2k} H_{01}\left(z^{-2}\right) + z^{-2k-1} H_{00}\left(z^{-2}\right).$$

Since $\det[\boldsymbol{H_p}(z)]$ is symmetric and must also contain delays to be a perfect reconstruction system, for k=0 we obtain

$$\boldsymbol{H_p}(z) = \begin{bmatrix} H_{00}(z) & H_{01}(z) \\ -H_{01}(z^{-1}) & H_{00}(z^{-1}) \end{bmatrix}, \tag{A.47}$$

and, therefore,

$$\det[\boldsymbol{H_p}(z)] = H_{00}(z) H_{00}\left(z^{-1}\right) + H_{01}(z) H_{01}\left(z^{-1}\right) = 1, \tag{A.48}$$

which corresponds to the poly-phase equivalent of (A.44). On the unit circle $z = e^{i\omega}$, this is equivalent to

$$\left|H_{00}\left(e^{i\omega}\right)\right|^2 + \left|H_{01}\left(e^{i\omega}\right)\right|^2 = 1.$$

Thus, the necessary form (A.46) also means that

$$z^{-2k-1}\left[H_0(z) H_0\left(z^{-1}\right) + H_0(-z) H_0\left(-z^{-1}\right)\right] = 2z^{-2k-1},$$

or equivalently, on the unit circle

$$\left|H_0\left(e^{i\omega}\right)\right|^2 + \left|H_0\left(e^{i\omega+\pi}\right)\right|^2 = 2. \tag{A.49}$$

In other words, $\boldsymbol{H_p}(z)$ is unitary on the unit circle. Note that the product $\left[\boldsymbol{H_p}(z^{-1})\right]^{\mathrm{T}} \boldsymbol{H_p}(z)$ corresponds to a 2×2 identity matrix. Such matrix is called paraunitary matrix (Vetterli 1986, Vaidyanathan 1990).

The above discussion indicates two possible filter design approaches for orthogonal implementations. The first consists in finding a symmetric autocorrelation function with a single even-indexed coefficient (the central one) different from zero (Smith and Barnwell 1984, Mintzer 1985). The zeros on the unit circle of its square roots, $H_0(z)$ and $H_0(z^{-1})$, have to be double in order to prevent the function from changing sign. The second method synthesizes paraunitary matrices by means of lattice structures from which factorizations have been given by Vaidyanatan and Hoang (1988). Once the poly-phase matrix is obtained, the filters follow easily from (A.39).

A.2.3.2 Biorthogonal or General Perfect Reconstruction Filter Banks

In the biorthogonal case, the wavelet transform is invertible but not necessarily orthogonal. The main reason is that giving up orthogonality allows the filter design to be more flexible, making possible the development of filters of arbitrary size and symmetry while maintaining the the linear phase property. This is preferred in order to increase compression efficiency, since the shape of the basis functions can then be modeled to fit more specific structures.

From equation (A.42) we know that perfect reconstruction requires $\det\left[\boldsymbol{H_m}(z)\right]$ to be an odd delay, and the synthesis filters are $C(z) = cz^{-l}$. Thus for $l = 1$, it can be shown that $G_0(z) H_1(z)$ and $G_1(z) H_0(z)$ have only odd coefficients, that is,

$$\left\langle \overline{g}_0[n-2k], h_1[n-2l] \right\rangle = 0, \tag{A.50}$$

$$\left\langle \overline{g}_1[n-2k], h_0[n-2l] \right\rangle = 0, \tag{A.51}$$

where $\overline{x}[n] = x[-n]$. In matrix notation,

$$\boldsymbol{H_0 G_1} = 0 = \boldsymbol{H_1 G_0} \tag{A.52}$$

where $\boldsymbol{H_i}$ and $\boldsymbol{G_i}$ are defined in a way similar to (A.9). Since it is a perfect reconstruction system, the product $G_0(z) H_0(z) = z^{-1} H_1(-z) H_0(z)$ has a single nonzero odd coefficient. Similarly, the product $G_1(z) H_1(z) = -z^{-1} H_0(-z) H_1(z)$ has a single even-indexed coefficient. Thus,

$$\langle \overline{g}_i [n-2k], h_i [n-2l] \rangle = \delta_{kl}. \tag{A.53}$$

In operator notation, that is equivalent to

$$\boldsymbol{H_0 G_0} = \boldsymbol{I} = \boldsymbol{H_1 G_1}, \tag{A.54}$$

$$\boldsymbol{G_0 H_0} + \boldsymbol{H_1 G_1} = \boldsymbol{I}. \tag{A.55}$$

The coefficients of the low-pass filter are unrelated to the high-pass one, requiring therefore the use of an additional basis set. Nevertheless, some relation of orthogonality must exist between the analysis and synthesis filter if perfect reconstruction is to be obtained within tree-structured transforms. Note that although $\boldsymbol{H_0 G_0}$and $\boldsymbol{H_1 G_1}$ are not necessarily orthogonal, orthogonality exists through (A.50) (A.51) and (A.53) and, hence, the name biorthogonal. In the special case where we have a paraunitary solution it holds that $\boldsymbol{G_0} = \boldsymbol{H_0^*}$ and $\boldsymbol{G_1} = \boldsymbol{H_1^*}$, leading to projections onto subspaces which are mutually orthogonal. The two most frequent classes of biorthogonal filters with linear phase are the following:

1. *Both filters, $H_0(z)$ and $H_1(z)$, are symmetric and of odd lengths, differing by an odd multiple of 2.*

2. *One of the filters is symmetric and the other antisymmetric; both lengths are even although they can be either equal or differing by a multiple or 2.*

Since the synthesis filters are $G_0(z) = H_1(-z)$ and $G_1(z) = -H_0(z)$, in the first case we use $g_0[n] = (-1)^n h_1[n]$ for the low-pass filter, and $g_1[n] = (-1)^{n+1} h_0[n]$ the high-pass filter, whereas in the second case both synthesis filters alternate sign and, thus, $g_0[n] = (-1)^{n+1} h_1[n]$ and $g_1[n] = (-1)^n h_0[n]$, respectively.

Now we can show that infinitely iterated biorthogonal perfect reconstruction filter-banks generate biorthogonal sets of functions. Let us denote as $H_0^{(j)}(z)$ and $G_0^{(j)}(z)$ the filters which are equivalent to the cascade of j blocks of filtering/sub-sampling in the analysis and synthesis sections, but respectively. These filters are assumed to be regular but each one defines a different piecewise constant function on the interval $1/2^j$,

$$f^{(j)}(t) = 2^{j/2} h_0^{(j)}[n] \qquad n/2^j \leq t \leq (n+1)/2^j \tag{A.56}$$

$$\widetilde{f}^{(j)}(t) = 2^{j/2} \overline{g}_0^{(j)}[n] \qquad n/2^j \leq t \leq (n+1)/2^j. \tag{A.57}$$

Since we are assuming regularity, both $f^{(j)}(t)$ and $\widetilde{f}^{(j)}(t)$ converge to continuous functions as $j \to \infty$. It can be shown, following the same analytical scheme of the preceding subsection that these functions satisfy the two-scale difference equations,

$$\phi(t) = 2^{1/2} \sum_{n=0}^{L-1} h_0[n]\, \phi(2t-n) \tag{A.58}$$

$$\widetilde{\phi}(t) = 2^{1/2} \sum_{n=0}^{L-1} \overline{g}_0[n]\, \widetilde{\phi}(2t-n). \tag{A.59}$$

Similarly, the associated bandpass functions are defined by

$$\psi(t) = 2^{1/2} \sum_{n=0}^{L-1} h_1[n]\, \phi(2t-n) \tag{A.60}$$

$$\widetilde{\psi}(t) = 2^{1/2} \sum_{n=0}^{L-1} \overline{g}_1[n]\, \widetilde{\phi}(2t-n). \tag{A.61}$$

It can be shown inductively that $\phi(t)$ and $\widetilde{\phi}(t)$ are orthogonal with respect to integer shifts. At the 0th level, $\widetilde{f}^{(0)}(t)$ and $f^{(0)}(t)$ are each equal to the indicator function on the interval [0,1), so $\left\langle \widetilde{f}^{(0)}(t-l), f^{(0)}(t-k) \right\rangle = \delta_{kl}$. Assuming orthogonality at the jth level, $\left\langle \widetilde{f}^{(j)}(t-l), f^{(j)}(t-k) \right\rangle = \delta_{kl}$, it holds that

$$\begin{aligned}
&\left\langle \widetilde{f}^{(j+1)}(t-l), f^{(j+1)}(t-k) \right\rangle \\
&\qquad = 2 \left\langle \sum_n \overline{g}_0[n]\, \widetilde{f}^{(j)}(2t-2l-n), \right. \\
&\qquad\quad \left. \sum_m h_0[m]\, f^{(j)}(2t-2k-m) \right\rangle \\
&\qquad = \sum_n \overline{g}_0[n]\, h_0[n+2l-2k] \\
&\qquad = \delta_{kl}.
\end{aligned}$$

Therefore, in the limit we obtain

$$\left\langle \widetilde{\phi}(t-l), \phi(t-k) \right\rangle = \delta_{kl}. \tag{A.62}$$

Once this is established, the relation

$$\left\langle \widetilde{\psi}\left(t-l\right), \psi\left(t-k\right)\right\rangle = \delta_{kl} \tag{A.63}$$

follows immediately from (A.53), whereas the relations

$$\left\langle \widetilde{\phi}\left[n-2k\right], \psi\left[n-2l\right]\right\rangle = 0, \tag{A.64}$$

$$\left\langle \widetilde{\psi}\left[n-2k\right], \phi\left[n-2l\right]\right\rangle = 0, \tag{A.65}$$

come from (A.50) and (A.51), respectively. Therefore, the conditions for perfect reconstruction on the filter coefficients lead to functions with biorthogonality properties as shown above.

Orthogonality across scales is also verified following the analysis of subsection A.2.2.1:

$$\left\langle 2^{-j/2}\widetilde{\psi}\left(2^{j}t-l\right), 2^{-i/2}\psi\left(2^{-i}t-k\right)\right\rangle = \delta_{ij}\delta_{kl}. \tag{A.66}$$

Thus the set $\left\{\psi\left(2^{j}t-l\right), \widetilde{\psi}\left(2^{i}t-l\right); i,j,k,l\in\mathbb{Z}\right\}$ is biorthogonal. Completeness can be verified as in the orthogonal case in (Cohen et al. 1992). Hence any function in $L^2\mathbb{R}$ can be written

$$x\left(t\right) = \sum_{j}\sum_{k}\left\langle 2^{-j/2}\psi\left(2^{j}t-k\right), x\left(t\right)\right\rangle 2^{-j/2}\widetilde{\psi}\left(2^{j}t-k\right).$$

Additionally, note that $\psi\left(t\right)$ and $\widetilde{\psi}\left(t\right)$ play interchangeable roles. So, regular biorthogonal FIR filters lead to biorthogonal bases of functions of finite length. The converse is also true and is shown briefly. Assume that we have functions $\psi\left(t\right)\phi\left(t\right)\widetilde{\phi}\left(t\right)\widetilde{\psi}\left(t\right)$ satisfying (A.58) - (A.61) and (A.62) - (A.65). Then, it can be verified that they can be used to generate filter-banks. For example, using (A.62)

$$\begin{aligned}
&\left\langle \widetilde{\phi}\left(t-l\right), \phi^{(j+1)}\left(t-k\right)\right\rangle \\
&\qquad = \left\langle \sum_{n}\overline{g}_0\left[n\right]\widetilde{\phi}^{(j)}\left(2t-2l-n\right), \right. \\
&\qquad\quad \left. \sum_{m}h_0\left[m\right]\phi^{(j)}\left(2t-2k-m\right)\right\rangle \\
&\qquad = \sum_{n}\sum_{m}\overline{g}_0\left[n\right]h_0\left[m\right]\left\langle \widetilde{\phi}\left[2t-2l-n\right], \right. \\
&\qquad\quad \left. \phi\left[2t-2k-m\right]\right\rangle \\
&\qquad = \sum_{n}\overline{g}_0\left[n\right]h_0\left[n+2l-2k\right] = \delta_{kl}.
\end{aligned}$$

The other filter biorthogonality relations follow from (A.63) - (A.65).

A.3 Filter Design

In the preceding section we showed that the orthonormal scaling function $\phi(t) = \sin(\pi t)/\pi t$ is not useful because it assumes the use of a low-pass filter of infinite length, $h_0[n] = 1/\sqrt{2}\phi(n/2)$ for $n \to \infty$, impossible to handle in practice. A possible solution to this problem can be the interpolation of "spline functions" up to some degree p. For example, if we rewrite Eq. (A.28) in the frequency domain and use the Fourier transform

$$\dot{\phi}\left(\frac{\omega}{2\pi}\right) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^{p+1}, \tag{A.67}$$

solving for $h_0[n]$ gives the binomial filter

$$h_0[n] = 2^{-(p+1)} \begin{pmatrix} p+1 \\ n \end{pmatrix}, \quad \text{for} \quad n = 0, \ldots, p+1, \tag{A.68}$$

and the corresponding synthesis filter $g_0[n] = (-1)^{n+1} h_0[L-1-n]$ then follows immediately.

In general, remark B1 ensures point-wise convergence to a continuous function that will not necessarily be regular. Ideally, to get a maximally regular filter, written as $H(z) = \left[(1+z)^{-1}/2^N\right] F(z)$ (see section A.2.2.1), we would maximize N while simultaneously minimizing $\sup_{\omega\in[0,2\pi]} |F(e^{i\omega})|$, but only the parameter N can be controlled easily. The question of irregularity is more related to the biorthogonal case where obtaining such characteristics is quite difficult. Note that the regularity of both analysis and synthesis filters has to be checked. For example in the 4 length linear phase case, while the low-pass impulse response (which is of the form$[1, \alpha, \alpha, 1]$) is regular for large positive values of α, the vector $[1, -\alpha, -\alpha, 1]$ (which corresponds to $H_1(-z)$) is irregular (see Fig. A.3).

From subsection A.2.3 it can be seen using (A.37), (A.40) and (A.41) that the condition to ensure perfect reconstruction with FIR synthesis filters, after an FIR analysis section, can be expressed in two forms:

$$H_{00}(z) H_{11}(z) - H_{01}(z) H_{10}(z) = z^{-l} \tag{A.69}$$

$$H_0(z) H_1(-z) - H_0(-z) H_1(z) = 2z^{-2l-1}. \tag{A.70}$$

One equation implies the other and both can be used in different ways to design the filters, depending on what is easier for us. In particular, they are special forms of what is known as the Bezout identity equation (Kailath 1980, Schroeder 1986):

$$a(x)\,p(x) + b(x)\,q(x) = c(x)\,. \tag{A.71}$$

When two of the polynomials are known, i.e. $a(x)$ and $b(x)$, it can be shown that a solution for $[p(x), q(x)]$ exists if the greatest common divisor, $\gcd[a(x), b(x)]$, divides $c(x)$. This implication introduces the following constraints on the filter-banks which are generally used within the design process.

Remark. **B.2:** *Assume that the filters $H_0(z)$ and $H_1(z)$ are FIR and causal.*

1. *Then, given one of the pairs $[H_{00}(z), H_{01}(z)]$, $[H_{00}(z), H_{10}(z)]$, $[H_{11}(z), H_{01}(z)]$ or $[H_{11}(z), H_{10}(z)]$ in order to achieve perfect reconstruction with the other pair it is necessary and sufficient that the given pair be co-prime (except for possible zeros at $z = \infty$).*

2. *A filter $H_0(z)$ has a complementary filter if and only if it has no zeros in pairs at $z = \alpha$ and $z = -\alpha$.*

3. *There is always a complementary filter to the binomial filter:*

$$H_0(z) = \left(1 + z^{-1}\right)^{-k} = H_{00}\left(z^2\right) + z^{-1} H_{01}\left(z^2\right). \tag{A.72}$$

That the poly-phase pair must be co-prime is easily obtained by using the previous Benzout condition in (A.69). Since the gcd must divide the right hand of (A.69), the only factors they can have in common are zeros at $z = \infty$. For the same reason, it should also be clear that for $H_0(z)$ and $H_1(z)$ to form a perfect reconstruction pair it is necessary that they be co-prime.

On the other hand, $H_0(z)$ has a zero pair at $(-\alpha, \alpha)$ if and only if it has a factor $A(z^2)$. This can only happen if both $H_{00}(z)$ and $H_{01}(z)$ contain a common polynomial $A(z)$ (they are not co-prime), but this contradicts the first point of the remark. Thus, $H_0(z)$ cannot have zeros of the form $(-\alpha, \alpha)$. Finally, if $H_{00}(z)$ and $H_{01}(z)$ had a common factor it would appear as a pair of zeros of $H_0(z)$ at $(-\alpha, \alpha)$ but $H_0(z)$ only has zeros at z=-1 when given as (A.72), so they cannot have such a factor. Thus, if the low-pass filter takes the form $H_0(z) = (1 + z^{-1})^{-k}$ a complementary filter $H_1(z)$ exists.

In practice, these conditions serve to determine whether a complementary filter exists in relation to some chosen $H_0(z)$. In the orthogonal case, the implications are tremendous since

these limitations are added to the fact that the reconstruction matrices are forced to be the transpose of the analytical ones. Additionally, since zeros are placed on the unit circle and coefficient vectors are normalized, it is preferable to prevent coefficient values not exceeding from 1 in absolute value.

For these reasons, a common choice is the template

$$\boldsymbol{H} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

which is generally used to develop low-pass and high-pass filters. In this case, the generation of perfect reconstruction orthogonal filters can be automated easily by means of the following *poly-phase matrix*

$$\begin{pmatrix} H_{00}(z) & H_{01}(z) \\ H_{10}(z) & H_{11}(z) \end{pmatrix} = \left(\prod_{k=0}^{K-1} \begin{pmatrix} \cos\vartheta_k & \sin\vartheta_k \\ -\sin\vartheta_k & \cos\vartheta_k \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \right) \times \begin{pmatrix} \cos\vartheta_K & \sin\vartheta_K \\ -\sin\vartheta_K & \cos\vartheta_K \end{pmatrix},$$

where $\{\vartheta = (\vartheta_0, \ldots, \vartheta_{K-1}) : \vartheta_k \in [0, \pi)\}$ are angle values. The angle values are introduced as inputs of the algorithm and initialize calculations. All them are chosen within $[0, \pi)$, except for ϑ_K which is the residue $\pi/4 - \sum_{k=0}^{K-1} \vartheta_k$ modulo 2π and can be within $[0, 2\pi)$. At first, the number of available filters can be considerable depending on the order of the filter and the discrete basis used for ϑ, $\left\{ \frac{\pi}{N}(t_0, \ldots, t_{K-1}) : t_l \in \{0, \ldots, N-1\} \right\}$ (in fact, many of the available standard orthogonal filters; Daubechies, Coiflets, Symlets and so on; can be obtained from this approach). However, this behavior does not involve a significant improvement compared to biorthogonal filters since having coefficient values restricted within $[-1, 1)$ does not filter with good compression capabilities. After the poly-phase matrix $\boldsymbol{H_p}(z)$ has been obtained, the low-pass and high-pass filters follows directly from expression (A.39).

A.3.1 Some Notes on Biorthogonal Filter Development

Remark B2 establishes the general conditions under which, given a filter $H_0(z)$, a complementary filter $H_1(z)$ exists. In this case, the Euclidean algorithm (Cohen 1993) serve as a verification method for such filter (more precisely, the result from $\gcd[H_0(z), H_1(z)]$ must divide the right-hand of A.70). For this reason, first the poly-phase matrix $\boldsymbol{H_p}(z)$, is obtained and then expression A.39 is used to calculate the filters.

However, one can design filters that take the form of A.72 in order to ensure a certain degree of regularity. In this sense, it can be shown that the poly-phase components of $H_0(z)$ and $H_1(z)$ generated by the lattice structure

$$\boldsymbol{H_p}(z) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \prod_{i=1}^{N} \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix} \begin{bmatrix} 1 & \alpha_i \\ \alpha_i & 1 \end{bmatrix}, \tag{A.73}$$

leads to $H_0(z) = (1+z^{-1})^{2N+1}$. Here, the coefficients $(\alpha_1, \alpha_2, \ldots, \alpha_N)$ are chosen as

$$\alpha_i = \frac{2N + (-1)^{i+1}}{2(N+i) - 1}. \tag{A.74}$$

The complement of $H_0(z)$ is not unique, so an important issue arises in order to find them. In practice, we specifically design $H_0(z)$ to meet some application requirements and then determine the complement that is needed to complete the perfect reconstruction filter-bank.

In this section, we address the design of complementary filters to the low-pass form $H_0(z) = (1+z^{-1})^{2k}$. The main advantage lies on an expression for $P(z)$, similar to A.24, which implicitly sets a regularity constraint for the filters (Vetterli 1986). We first focus on linear phase solutions of odd length (first type of biorthogonal filters, see subsection A.2.3.2) and then we extend to other more general solutions, providing some mechanisms for obtaining other complementary filters when one is available. Finally, a brief discussion about the verification of regularity is provided.

A.3.1.1 Linear Phase Filter Design with a Maximum Number of Zeros at π

The present method is based on the assumption that having the maximum number of zeros at z = -1 adds regularity to the filters (see section A.2.2). For this purpose, we write $P(z) = (1+z^{-1})^{2k} R_{2k}(z)$ and we determine a complementary filter $R_{2k}(-z)$ such that condition (A.70) is satisfied. Note that by using $(1+z^{-1})^{2k}$ we are assured that a complementary filter to the binomial of any degree exists (see remark B.2.3). However, if $P(z)$ has even a single zero at z = -1, it cannot have any at $z = 1$ and vice-versa (see remark B.2.2). Also recall that $P(z)$ must have a single non-zero odd-indexed coefficient (normalized to 1) for the filter-bank to be a perfect reconstruction pair.

Therefore, we wish to find an $R_{2k}(z)$ so that $P(z)$ is valid, as defined by (A.43). Since $R_{2k}(z)$ is written as

$$R_{2k}(z) = r_0 + r_1 z^{-1} + \ldots r_{k-1} z^{-k+1} + \ldots + r_0 z^{-2(k-1)}, \tag{A.75}$$

to be symmetric, on equating to zero the terms of z of odd power (except for the central one, $r_{k-1}z^{-k+1}$, which must be one) we obtain a linear equation system of size k. So to obtain the missing coefficients $\boldsymbol{r}_{2k} = (r_0, \ldots, r_{k-1})$ we solve

$$\boldsymbol{F}_{2k} \cdot \boldsymbol{r}_{2k} = \boldsymbol{e}_{2k} \tag{A.76}$$

where $\boldsymbol{F}_{2k}$ is the k by k matrix and $\boldsymbol{e}_{2k}$ is the length k column vector $(0;0;\ldots;1)$. For example, if we calculate $P(z) = (1+z^{-1})^6 R_6(z)$, for k=3 we obtain

	z^0	z^{-1}	z^{-2}	z^{-3}	z^{-4}	z^{-5}	z^{-6}	z^{-7}	z^{-8}	z^{-9}	z^{-10}
r_0	1	6	15	20	15	6	1				
r_1		1	6	15	20	15	6	1			
r_2			1	6	15	20	15	6	1		
r_1				1	6	15	20	15	6	1	
r_0					1	6	15	20	15	6	1

,

so imposing the constraints on the coefficients of the odd powers of z^{-1} the equation system becomes

$$\begin{pmatrix} 6 & 1 & 0 \\ 20 & 16 & 6 \\ 12 & 30 & 20 \end{pmatrix} \begin{pmatrix} r_0 \\ r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

giving $\boldsymbol{r_6} = (3/2, -9, 19)/128$.

Having found $R_{2k}(z)$, its factors can be regrouped with the *2k* zeros at z =- 1 to form the two filters $H_0(z)$ and $H_1(-z)$, both of which are to be regular. For a linear filter of odd length N, the $N-2$ length complementary solution will be unique since it leads to a system of $(N-1)/2$ equations in $(N-1)/2$ unknowns. Similarly, the solution for a filter of even length N will be unique but will have the same length.

A.3.1.2 *Diophantine* Equations and Complementary Filters

Again, we make use of the interpretation of condition A.70 as a Benzout identity to determine other complementary filters to $H_0(z)$, when one complementary filter is available: for example, using the method shown above. In this sense, if solutions of the form

$$\boldsymbol{a}\boldsymbol{p}' + \boldsymbol{b}\boldsymbol{q}' = 0, \tag{A.77}$$

exist they can be added to $(\boldsymbol{p},\boldsymbol{q})$ and generate new solutions to (A.71). This fact is exploited here by noting that polynomials analogous to (A.77) can be found easily. They are known as Diophantine equations. In fact, if we work with the modulation version of (A.70) and identify $\boldsymbol{a} = -H_0(z)$ and $\boldsymbol{b} = H_0(-z)$ it can be shown that using $[\boldsymbol{p},\boldsymbol{q}] = [E(z)H_0(-z), E(-z)H_0(z)]$ sets the right-hand to zero whenever $E(z) = E(-z)$.

Remark. **B.3**: *Given a linear phase filter $H_0(z)$ of odd length N, its length N-2 linear phase complement $H_1(z)$, and the template*

$$E(z) = \sum_{i=1}^{m} \alpha_i \left(z^{-2(i-1)} + z^{-(4m-2i)} \right), \tag{A.78}$$

all higher degree odd length linear phase filters complementary to $H_0(z)$ are of the form

$$H_1^{'}(z) = z^{-2m} H_1(z) + E(z) H_0(z). \tag{A.79}$$

Note that $E(z) = E(-z)$ and $E(z) H_0(z)$ is symmetric about the point $(N{+}4m{-}3)/2$, just as $z^{-2m} H_1(z)$ is. Hence, $H_1^{'}(z)$ is a linear phase solution of length $N + 4m - 2$. Since that would increase the length of $H_0(z) H_1^{'}(-z)$ to $2N + 4m - 3$, but the properties of $P(z)$ must be kept intact, the new polynomial must be of the form $P^{''}(z) = z^{-2m} H_0(z) H_1(-z)$ to be valid (with length $2N - 3$ and odd-indexed coefficients set to zero, except the central one). Therefore, the constants $(\alpha_1, \ldots, \alpha_m)$ need to be configured in such a way that $4m$ of the end terms in $P^{''}(z)$ are canceled. This is possible thanks to the selected form in (A.78), which leads to an expression $H_1^{'}(-z) = z^{-2m} H_1(-z) + E(z) H_0(-z)$ equivalent to (A.79). Since $H_0(z)$ is a common factor and solution and $H_1(-z)$ is unique, all length $N + 4m - 2$ solutions will be of this form.

As an example, consider the filter $H_0(z) = [1, 4, 6, 4, 1]$ and its unique 3-length filter $H_1(z) = -[1, 4, 1]/16$. If $m = 2$, we obtain

	z^0	z^{-1}	z^{-2}	z^{-3}	z^{-4}	z^{-5}	z^{-6}	z^{-7}	z^{-8}	z^{-9}	z^{-10}
$\alpha_1 H_0(z)$	α_1	$4\alpha_1$	$6\alpha_1$	$4\alpha_1$	α_1						
$z^{-2}\alpha_2 H_0(z)$			α_2	$4\alpha_2$	$6\alpha_2$	$4\alpha_2$	α_2				
$z^{-4}\alpha_2 H_0(z)$					α_2	$4\alpha_2$	$6\alpha_2$	$4\alpha_2$	α_2		
$z^{-6}\alpha_1 H_0(z)$							α_1	$4\alpha_1$	$6\alpha_1$	$4\alpha_1$	α_1
$z^{-4} H_1(z)$					$-\frac{1}{16}$	$-\frac{1}{4}$	$-\frac{1}{16}$				

Giving for $H_1^{'}(z)$

$$H_1^{'}(z) = \left[\alpha_1, 4\alpha_1, 6\alpha_1 + \alpha_2, 4(\alpha_1 + \alpha_2), -\frac{1}{16} + \alpha_1 + 7\alpha_2, 8\alpha_2 - \frac{1}{4}, -\frac{1}{16} + \alpha_1 + 7, 4(\alpha_1 + \alpha_2), 6\alpha_1 + \alpha_2, 4\alpha_1, \alpha_1 \right],$$

which is linear phase and complementary to $H_0(z)$, for some α_i.

An extended result allows the use of Diophantine equations to obtain more general solutions. This is possible by noting that any length N filter has at most $N - 2$ complementary filters of length $N - 2$ (since the length of $P(z)$ is of length $2N - 3$, from which $N - 2$ are

odd-indexed, the single nonzero odd-indexed coefficient can be placed in any of these $N-2$ positions, leading to $N-2$ equation systems of size $N-2$).

Remark. **B.4**: *All filters of length $N+2m-2$, complementary to a length N length filter $H_0(z)$, have the form*

$$H_1'(z) = z^{-2k} H_1(z) + E(z) H_0(z) \tag{A.80}$$

where $E(z) = E(-z)$ is a polynomial of the form

$$E(z) = \sum_{i=0}^{m-1} \alpha_i z^{-2i}, \tag{A.81}$$

$k \in \{0, 1, \ldots, m\}$ and $H_1(z)$ is a length $N-2$ complementary filter.

That this is a solution is verified by direct substitution of $E(z)$, which reduces the length of $P''(z) = H_0(z)\left[H_1'(-z) - E(z) H_0(-z)\right]$ by $2m$ for some choice of α_i. Since $H_0(z)$ is still a factor of both $P(z)$ and $P''(z)$, the remaining factor must be one of the $N-2$ length solutions, for which there are only $N-2$.

A.3.1.3 Discussion

While the design procedure of the previous methods are fairly simple, the spectra of the scaling function $\dot{\phi}(\omega)$ and the wavelet $\dot{\psi}(\omega)$, may not be as one might wish from low-pass and band-pass filters. In the first method, it turns out that ensuring that both $H_0(z)$ and $H_1(-z)$ meet the condition of remark B.1 can force one to choose filters of quite unequal length when k is small, and only when k increases does it becomes possible to obtain filters of quite regular wavelets.

However, other methods aimed at optimizing regularity permit more robust designs even when failing in the verification of this condition. Another possible approach is to note that there is no need to place all of the zeros at $z=-1$ to start with. For example, we could calculate the complementary filter to a factor $(1+z^{-1})^{2k}\left[u_1(z), u_2(z), \ldots, u_j(z)\right]^2$ where $u_i(z)$ represents a zero pair on the unit circle. We are then assured of having a factor $(1+z^{-1})^{k}\left[u_1(z), u_2(z), \ldots, u_j(z)\right]$ to place in the stop band of each of the filters.

Although higher degree solutions can give better results, in general FIR filters lead to very constrained solutions. Further methods on wavelet design can be found, for example, in Strang and Nguyen (1996), Chui (1992) and Mallat (1990). If compact support is not desired, similar techniques using IIR filters generate orthogonal wavelets (Herley and Vetterli 1991, 1992).

Appendix B

CWT Computation with the DWT

Full discretization of the scale and time parameters is addressed by setting

$$s = a_0^j, \tag{B.1}$$

$$u = k, \tag{B.2}$$

where $1 \leq a_0 \leq 2$ and $k = 1, \ldots, N$. However in a more practical implementation, full discretization of the time-frequency plane is developed by maintaining a dyadic discretization in one parameter, while the other takes denser values in the Wavelet Series approach. The general case, which leads to the CWT computation, is obtained by combining both techniques.

B.1 Finner Sampling in Scale

Here, we maintain $u = k2^j$, while the scale parameter is sampled according to

$$s = 2^{j+m/M}, \quad \text{for} \quad m = 0, \ldots, M-1 \tag{B.3}$$

where m is called the "voice".

The simplest way to compute WS coefficients on M voices per octave consists in exploiting the standard octave-by-octave algorithm as the main building function block. That is, for each m replace $\psi(t)$ by the slightly stretched wavelets $2^{-m/2M}\psi\left(2^{-(j+m/M)}t\right)$ in Eq. 3.14. Similarly, replace $\phi(t)$ by $2^{-m/2M}\phi\left(2^{-(j+m/M)}t\right)$. Both scaling and wavelets basis functions become

$$2^{-(j+m/M)/2}\phi\left(2^{-(j+m/M)}t\right) \quad \text{and} \quad 2^{-(j+m/M)/2}\psi\left(2^{-(j+m/M)}t\right) \tag{B.4}$$

which are used respectively in Eq. A.28 and Eq. A.29 to obtain m pairs of filters ($\boldsymbol{h}_0^{(m)}$ and $\boldsymbol{h}_1^{(m)}$) corresponding to each voice m. Thus, the whole algorithm requires M times the computational load of the octave-by-octave algorithm defined in Eq. 3.35.

$$\hat{C}\left[j_m, k\right] = \mathrm{CWT}\left\{x\left(t\right); 2^{j_m}, k2^{j_m}\right\} = \mathrm{DWT}_m\left\{\hat{x}\left[n\right], 2^{j_m}; k2^{j_m}\right\}, \tag{B.5}$$

where $j_m = j + m/M$ for $j_m \in \mathbb{R}$ and $j \in \mathbb{Z}$. Thus, initializing $a_{j_m} = \hat{x}\left[n\right]$ for $j = 0$ and all $m = 1, \ldots, M$; at decomposition we have

$$a_{j_m+1}\left[l\right] = \boldsymbol{a}_{j_m} * \overline{\boldsymbol{h}}_0^{(m)}\left[2l\right] \quad \text{and} \quad c_{j_m+1}\left[l\right] = \boldsymbol{a}_{j_m} * \overline{\boldsymbol{h}}_1^{(m)}\left[2l\right] \tag{B.6}$$

B.2 Finner Sampling in Time: The "à trous" Algorithm

Here, we restrict the scale to an octave-by-octave computation $s = 2^j$, while considering all possible values for the time parameter $u{=}k$. Note that the WS coefficients are nothing but part of the computation required here

$$\hat{C}\left[j, k\right] = \mathrm{CWT}\left\{x\left(t\right); 2^j, k\right\} = \mathrm{DWT}\left\{\hat{x}\left[n\right]; 2^j, k\right\}, \tag{B.7}$$

with the difference, of course, that DWT now extends calculations to all integer values of u, instead of restricting to the values in $k2^j$.

For any filter $f\left[n\right]$, let $f^{(j)}\left[n\right]$ denote the filter sequence obtained by inserting $2^j - 1$ zeros between each sample of $f\left[n\right]$. This modification of the filters, equivalent to Eq. 3.25 and which creates holes in the sequence (*trous* in French), is very important since it avoids subsampling of the coefficient vector. Note that this is necessary to compute all the coefficients for $u = k$ while jumping from one octave to the other (Fig. B.1).

Then, at decomposition we have

$$a_{j+1}\left[l\right] = \boldsymbol{a}_j * \overline{\boldsymbol{h}}_0^{(j)}\left[l\right] \quad \text{and} \quad d_{j+1}\left[l\right] = \boldsymbol{a}_j * \overline{\boldsymbol{h}}_1^{(j)}\left[l\right], \tag{B.8}$$

and at reconstruction we have

$$a_j\left[l\right] = \frac{1}{2}\left(\boldsymbol{a}_{j+1} * \boldsymbol{g}_0^{(j)}\left[l\right] + \boldsymbol{d}_{j+1} * \boldsymbol{g}_1^{(j)}\left[l\right]\right). \tag{B.9}$$

As well as DWT, this algorithm also obtains a dyadic wavelet representation of the signal in a_0, but with node coefficients of length N which are calculated up to scale 2^J:

$$\left[\left\{\boldsymbol{c}_j\right\}_{1 \le j \le J}, \boldsymbol{a}_J\right]. \tag{B.10}$$

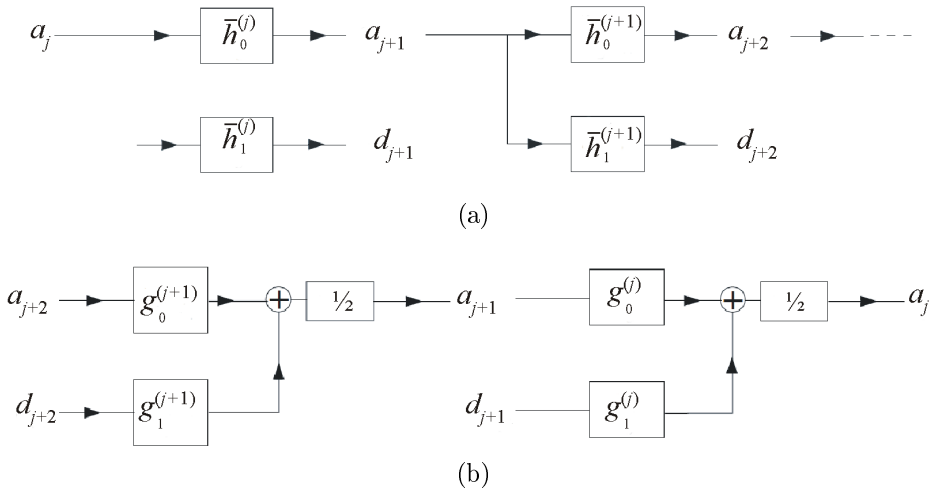


Figure B.1: Dyadic wavelet transform; (a) The coefficients are computed by cascading convolutions with dilated filters $\overline{\boldsymbol{h}}_0^{(j)}$ and $\overline{\boldsymbol{h}}_1^{(j)}$; (b) The original signal is reconstructed through convolutions with $\boldsymbol{g}_0^{(j)}$ and $\boldsymbol{g}_1^{(j)}$. A multiplication by $1/2$ is necessary to recover subsequent finer scales in $\boldsymbol{a}_j$.

This is why this method is sometimes referred as the fast "dyadic" wavelet transform. It is computed from $\boldsymbol{a}_0$ by cascading convolutions (B.8) for $0 \leq j \leq J$, as depicted in Fig. B.1a. The original signal $\boldsymbol{a}_0$ is recovered from its wavelet representation (B.10) by iterating (B.9) for $J \geq j \geq 0$, as illustrated in Fig. B.1b.

Appendix C

Multivariate Density Algorithms

C.1 Introduction

In chapter 5 we introduced the basic framework for density estimates in one-dimensional data. The extension of the AMISE calculation for the multivariate case is only slightly more complicated. The corresponding framework was provided by Scott (1992).

Given a sample $f(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^d$, the histogram is determined by regular partitions in the form of hyper-rectangles of size $h_1 \times h_2 \times \ldots \times h_d$. Consider a generic hyper-rectangular bin labeled G_m containing v_m points. As usual $\sum_m v_m = N$. Then, the multivariate histogram is defined by

$$\hat{f}(\mathbf{x}) = \frac{v_m}{N h_1 h_2 \cdots h_d}, \text{ for } \mathbf{x} \in G_m. \tag{C.1}$$

The asymptotically optimal bin widths, h_k^*, and resulting AMISE* for Normal multivariate data, $X \sim N(\mu, \Sigma)$ and $\Sigma = \text{Diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_d^2)$ are

$$\begin{aligned} h_k^* &= 2 \times 3^{1/(2+d)} \pi^{d/(4+2d)} \sigma_k N^{-1/(2+d)}, \\ \text{AMISE}^* &= 2^{-1/(1+d)} 3^{2/(2+d)} \pi^{-d^2/(4+2d)} c_d^{-1} N^{-2/(2+d)} \end{aligned} \tag{C.2}$$

where $c_d = \sigma_1 \sigma_2 \cdots \sigma_d$. In this case, since the constant in the bandwidth increases slowly from 3.4908 in one dimension to the limiting value $2\sqrt{\pi} = 3.545$ as $d \to \infty$, a very useful bin width expression is

$$h_k^* \approx 3.5 \sigma_k N^{-1/(2+d)}. \tag{C.3}$$

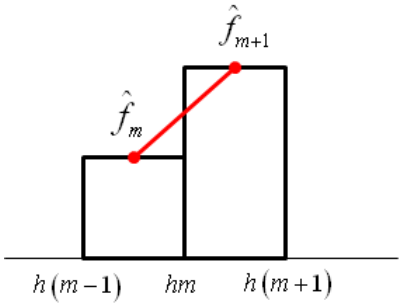


Figure C.1: The frequency polygon in a typical bin, $h\left(m-\frac{1}{2}\right) \leq x < h\left(m+\frac{1}{2}\right)$, which is derived from two adjacent histogram bins.

In the special bi-variate case, d=2, of normal data $f(x_1, x_2) = N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ where ρ denotes nonzero correlation between both distributions the optimal bin width and the AMISE can be rewritten as

$$\begin{aligned} h_i^* &= 3.504\sigma_i (1-\rho)^{3/8} N^{-1/4}, \\ \text{AMISE}^* &= \frac{0.122}{\sigma_1\sigma_2}\left(1-\rho^2\right)^{-3/4} N^{-1/2}. \end{aligned} \tag{C.4}$$

Thus, if the data are not independent but are clustering along a line, smaller bin widths are necessary to "track" this feature.

In general, the estimation is more accurate as long as Eq. C.1 becomes sharper (variance get smaller) when $N \to \infty$ and the region in $\mathbb{R}^d$, defined by the hyper-cube $V = h_1 h_2 \cdots h_d$, is so small that $\hat{f}(\mathbf{x})$ does not vary appreciably within it. Since N is fixed, to improve the accuracy of $\hat{f}(\mathbf{x})$ we could let V approach zero but then the region would become too small that it would enclose no samples, so we need to find a compromise value for V.

C.2 Kernel Density Estimation (KDE)

KDE methods and its variants: Frequency Polygons (FP) and Averaged Shifted Histograms (ASH); are preferable when the goal is to obtain robust estimates but with reduced datasets. In fact, FP is the natural choice after the histogram, since it connects the two adjacent histogram values, $\hat{f}_m$ and $\hat{f}_{m+1}$, between the bin centers as shown in Fig. C.1. For the uni variate case, that is

$$\hat{f}(x) = \left(\frac{1}{2} - \frac{x}{h}\right)\hat{f}_m + \left(\frac{1}{2} + \frac{x}{h}\right)\hat{f}_{m+1}, \quad \text{for } h\left(m-\frac{1}{2}\right) \leq x < h\left(m+\frac{1}{2}\right), \tag{C.5}$$

where $\hat{f}_i = v_i/(Nh)$.

A nearly equivalent and computationally efficient approach corresponds to the Averaged Shifted Histogram (ASH). Consider a collection of k histograms, $\hat{f}_1, \hat{f}_2, \ldots \hat{f}_k$, each with bin width h and computed from Eq. 5.7 but with bin origins

$$t_0 = 0, \frac{h}{k}, \frac{2h}{k}, \ldots, \frac{(k-1)\,h}{k},$$

respectively. Then, the uni-variate naive (or unweighted) ASH is defined as

$$\hat{f}_{\text{ASH}}(x) = \frac{1}{k}\sum_{i=1}^{k} \hat{f}_i(x). \tag{C.6}$$

Since the ASH is piecewise constant and of size $\triangle = h/k$, a more general expression of Eq. C.6 can be obtained if we refer the bin count v_m to the narrower interval $G_m \equiv [m\triangle, (m+1)\triangle]$. Indeed, the height of the ASH in G_m is the average of the heights of the k shifted histograms, each of width $h = k\triangle$:

$$\frac{v_{k+1-m} + \ldots + v_k}{Nh}, \frac{v_{k+2-m} + \ldots + v_{k+1}}{Nh}, \ldots, \frac{v_k + \ldots + v_{k+m-1}}{Nh}.$$

As such, the naive ASH is defined as

$$\begin{aligned} \hat{f}(x;k) &= \frac{1}{k}\sum_{i=1-k}^{k-1}\left(\frac{k-|i|}{Nh}\right) v_{m+i} \\ &= \frac{1}{Nh}\sum_{i=1-k}^{k-1}\left(1-\frac{|i|}{k}\right) v_{m+i} \quad \text{for} \quad x \in G_m. \end{aligned} \tag{C.7}$$

The weights on the bin counts in Eq. C.7 take on the shape of an isosceles triangle with base (-1,1). However, the parameter k and the bin origin t_0 should be omitted as far as possible, otherwise the ASH would be parametric.

With h and N fixed and k increasing, and taking into account the effect of a single data point x_j on the ASH estimate $\hat{f}(x)$ at an specific point x, the weight term is proportional to

$$1 - \frac{|i|}{k} = 1 - \frac{|i|\cdot\triangle}{k\cdot\triangle} = 1 - \frac{|x-x_j|}{h} + O(1/k), \tag{C.8}$$

if $|x - x_j| < h$ (since $x \in G_m$ and $x_j \in G_{m+i}$, the number of bins between x and x_j is approximately i and, hence, $|x - x_j| \approx |i|\cdot\triangle$). On the other hand, the influence will be zero if $x_j \notin [x-h, x+h)$.Thus, when $k \to \infty$, the general expression of the limiting ASH is obtained by means of Eq. C.7 as

Kernel	$K(x)$	*Kernel*	$K(x)$
Parzen window	$1/2$	Polynomial	$\frac{3}{4}\left(1-x^2\right)$,
Triangular	1-$\lvert x\rvert$		$\frac{15}{32}\left(1-x^2\right)\left(3-7x^2\right)$,
Epanechnikov	$\frac{3}{4}\left(1-\frac{x^2}{5}\right)/\sqrt{5}$		$\frac{105}{256}\left(1-x^2\right)\left(5-30x^2+33x^4\right)$,
Biweight	$\frac{15}{16}\left(1-x^2\right)^2$		$\frac{315}{4.096}\left(1-x^2\right)\left(35-385x^2+1.01x^4-715x^6\right)$
Triweight	$\frac{35}{32}\left(1-x^2\right)^3$	Cosine arch	$\frac{\pi}{4}\cos\left(\frac{\pi x}{2}\right)$
Gaussian	$\frac{1}{\sqrt{2}}\mathrm{e}^{-(1/2)x^2}$ for $\lvert x\rvert\leq\infty$	Skewed	$2.86\left(x+2/7\right)^3\left(5/7-x\right)^9$

Table C.1: Some common an unusual *Kernels*. All kernels are supported on [-1,1] unless noted otherwise

$$\lim_{k\to\infty}\hat{f}\left(x;k\right)=\frac{1}{Nh}\sum_{j=1}^{N}\left(1-\frac{\left|x-x_j\right|}{h}\right)I_{[-1,1)}\left(\frac{x-x_j}{h}\right), \tag{C.9}$$

where the sum is over the number of data points rather than the number of bins. In practice, the ASH is made continuous using the linear interpolation scheme provided by Eq. C.5 which is referred to as the frequency polygon of the averaged shifted histogram (FP-ASH).

Alternatively, Eq. C.9 can be written defining a *kernel function* $K(\cdot)$ of an isosceles triangle of density of the form,

$$K\left(t\right)=\left(1-\left|t\right|\right)I_{|-1,1)}\left(t\right). \tag{C.10}$$

The limiting ASH will then become,

$$\hat{f}\left(x\right)=\frac{1}{Nh}\sum_{i=1}^{N}K\left(\frac{x-x_i}{h}\right). \tag{C.11}$$

This formula (C.11) serves to define the *general Kernel Density Estimator* with kernel K. The ASH kernel (C.10) has always finite support, but most of the times an infinite-support kernel is chosen. Very often, the basic kernel estimator is written compactly as

$$\hat{f}\left(x\right)=\frac{1}{Nh}\sum_{i=1}^{N}K_h\left(x-x_i\right), \tag{C.12}$$

where $K_h\left(t\right)=K\left(t/h\right)/h$.

Graphically, the kernel method performs a density estimate by adding probability masses of size $1/n$, scaled by smoothing parameter h, at each data point (Fig. C.2). In practice, there are many different kernel choices that can be used for these tasks (see Table C.1). For a non-negative uni-variate kernel, Scott (1992) demonstrated the optimal AMISE* to be,

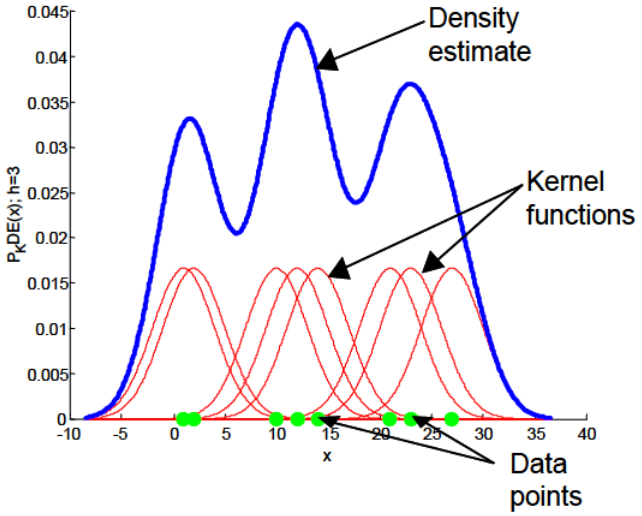


Figure C.2: Example of a Gaussian kernel estimate showing the individual kernels for each data point from (Osuna 2005). The kernel function determines the shapes of the bumps, and the "smoothing" parameter, h, determines their width.

$$\text{AMISE}^* = \frac{5}{4}\left[\sigma_K R\left(K\right)\right]^{4/5} R\left(f''\right)^{1/5} N^{-4/5}, \tag{C.13}$$

where $R(f")$ denotes roughness of the 2nd. kernel derivative, and σ_k its standard deviation. However, the quality of a density estimate is recognized to be primarily determined by the choice of smoothing parameter, and only in a minor way by the kernel choice.

Finally, analogous expressions can be inferred for the multivariate ASH and KDE, using the limiting ASH expression (C.11). Some algebra reveals that as $k_i \to \infty$, for $i = 1,...,d$; then

$$\hat{f}\left(\mathbf{x}\right) = \frac{1}{N h_1 h_2 \cdots h_d} \sum_{i=1}^{N} \left\{ \prod_{j=1}^{d} K\left(\frac{x_j - x_{i,j}}{h_j}\right) \right\}, \tag{C.14}$$

where K is the uni-variate isosceles triangle kernel (C.10) and (i, j) are used to index sample number and axis dimension, respectively. Kernel estimators are notoriously slow to compute, and some faster approximations should be considered. In this sense, the ASH is considered a *bona fide* density estimator since it provides a direct link to kernel methods. Therefore, it is a natural candidate for computation in practice.

C.3 The Nearest Neighbor Approach (k-NN)

The nearest neighbor density estimator represents an attempt to adapt the amount of smoothing to the local density of data. The degree of smoothing is controlled by an integer k, chosen to be considerably smaller than the sample size; typically $k \approx N^{-1/2}$.

Let the metric $d_i(x) = |x - x_i|$ be defined as the distance between point x and sample x_i, for each possible point x and arranged in ascending order

$$d_1(x) \leq d_2(x) \leq \ldots \leq d_N(x). \tag{C.15}$$

The *kth nearest neighbor density estimate* is then defined by

$$\hat{f}(x) = \frac{k}{2Nd_k(x)}. \tag{C.16}$$

To understand this definition, assume that the density at point x is $f(x)$. Then, if the size of the dataset is N, one would expect about $2Nhf(x)$ observations to fall within the interval [x-h,x+h], for each h>0. Since only k samples fall within the interval $[x - d_k(t), x + d_k(t)]$ by definition, an estimate of the density at point x may be obtained by setting $k = 2d_k(x) N\hat{f}(x)$.

$\hat{f}$ will always be positive and continuous, but will have discontinuous derivatives at all the same points d_k, creating all sort of spurious transitions and noise in the shape of $\hat{f}$ unless the value of k is relatively large. In contrast to KDE methods, the nearest neighbor estimate will not integrate to unity, making it unsuitable for normalization purposes. For x less than the smallest data point, we have $d_k(t) = x_{(k)} - x$ and for $x > x_{(N)}$ we will have $d_k(t) = x - x_{(N-k+1)}$. Using C.17, it follows that $\int_{-\infty}^{\infty} \hat{f}(x)\, dx = \infty$ and that the tails of $\hat{f}$ will die away at rate x^{-1}, extremely slowly.

The preceding concepts apply for the multivariate case extending the corresponding definitions to $\mathbf{x} \in \mathbb{R}^d$. Furthermore, it is possible to generalize the nearest neighbor estimate to provide a definition related to the kernel estimate. Let $K(t)$ be a kernel function integrating to one. Then, the *generalized kth nearest neighbor estimate* is defined by

$$\hat{f}(x) = \frac{1}{Nd_k(x)} \sum_{i=1}^{N} K\left(\frac{x - x_i}{d_k(x)}\right). \tag{C.17}$$

It can be seen at once that $\hat{f}(x)$ is precisely the kernel estimate evaluated at point x with window $h = d_k(x)$. Thus, the overall amount of smoothing is governed by the choice of the integer k, but the window width used at any particular point depends on the density of observations near that point.

As a final remark, it can be said that the number of related density methods is large and will not be reviewed further. There is the possibility to modify the kernel width from one data point to another in order to achieve adaptive smoothing (Jones 1990), increasing the accuracy of estimation even further at the expenses of a higher computational cost. Other approaches estimate densities from a quite different perspective. For example, *Orthogonal Series Approximations* (Kronmal and Tarter 1968, Watson 1969) are known to use kernels constructed from Fouriers basis functions. Other authors use optimization techniques (*Maximum penalized likelihood estimators*, MPL) to verify qualitative properties of density functions, such as the conflict between smoothness and goodness-of-fit (Delta methods, de Montricher et al. 1975, Good and Gaskings 1980, Klonias 1982).

Bibliography

P. Abry and P. Flandrin. On the initialization of the Discrete Wavelet Transform algorithm. *IEEE Sig. Proc. Letters*, 1:32–34, 1994.

M. Aguayo-Hernández. Biology and fisheries of chilean hakes (*merluccius gayi* and *merluccius australis)*. In *Hake: fisheries, ecology and markets*, pages 23–67. London. Chapman and Hall Book Series, 1955.

N. Ahmed and K.R. Rao. *Orthogonal Transforms for Digital Signal Processing*. Springer-Verlag, New York, 1975.

W.O. Ayres. Description of the new species of california fishes. In *Proceedings of California Academy of Natural Sciences*, volume 1, pages 23–67, 1855.

W. Banzhaf, P. Nordin, R. Keller, and F. Frankone. *Genetic Programming. An Introduction*. Morgan Kaufmann, 1998.

M. Baseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18:349–369, 1989.

T. Bayes. An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1764.

R.J. Beamish. Differences in the age of pacific hake (*merluccius productus*) using whole otoliths and sections of otoliths. *J. Fish. Res. Board Can.*, 36:141–151, 1979.

R.J. Beamish and D.A. Fournier. A method for comparing the precission of a set of age determinations. *Canadian Journal of Fisheries and Aquatic Sciences*, 38:982–983, 1981.

R.E. Bellman. *Adaptive Control Processes*. 1961.

A.P.M. Di Beneditto, R.M.A. Ramos, and N.R. Wille-Lima. *Os golfinhos: origem, classificaçao, captura acidental, hábito alimentar*. The Latin Americal Journal of Aquatic Mammals, 2001.

J.T. Bennett, G.W. Boehlert, and K.K. Turekian. Confirmation of longevity in *sebastes diplorproa* (pisces: Scorpaenidae) from pb-210/ra-2226 measurements in otoliths. *Marine Biology*, 71:209–215, 1982.

R.J.H. Beverton and S.J. Holt. *On the dynamics of exploited fish populations*. Chapman and Hall, London, UK, 1957.

G. Beylkin. Wavelets and fast numerical algorithms. In *Proceedings of Symposia in Applied Mathematics*, 1993.

J.L. Bird, D.T. Eppler, and D.M. Checkley. Comparison of herring otoliths using Fourier series shape analyses. *Canadian Journal of Fisheries and Aquatic Sciences*, 43:1228–1234, 1986.

C.M. Bishop. Probability density estimation. In *Neural Networks for Pattern Recognition*, pages 33–73. Oxford University Press, 1995.

R.W. Blacker. Recent advances in otolith studies. In *Sea Fisheries Research*, pages 67–90. Willey, 1995.

K.L. Bolles and G.A. Begg. Distinction between silver hake (*meluccius bilinearis*) stocks in u.s. waters of the North-west Atlantic based on otolith morphometrics. *Fish. Bull.*, 98: 451–462, 2000.

F.L. Bookstein. Tensor biometrics for changes in craneal shape. *Annals of Human Biology*, 11:413–437, 1984.

F.L. Bookstein. *Morphometric Tools for landmark data: Geometry and Biology*. Cambridge University Press, 1991.

F.L. Bookstein, R.E. Strauss, J.M. Humphries, B. Chernoff, R.L. Elder, and G.R. Smith. A comment upon the uses of Fourier methods in systematics. *Systematic Zoology*, 31:85–92, 1982.

B.E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the Fith Annual Workshop on Computational Learning.*, pages 144–152. ACM Press, 1992.

L. Botha. Growth and otolith morphology of the cape hakes *merluccius capensis* cast. and *merluccius paradoxus* franca. Technical report, Investigational Report of the Division of Sea Fisheries in South Africa, 1971.

L. Breiman, J.H. Friedman, R.A. olshen, and C.J. Stone. *Classification and Regression Trees.* Chapman & Hall, 1984.

G. Bruun. Z-transforms DFT filters and FFTs. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 26(1):56–63, 1978.

E.J. Burton, A.H. Andrews, H.K. Coale, and G.M. Cailliet. Application of radiometric age determination to three long-lived fishes using pb-210/ra-226 disequilibria in calcifieed structures: a review. *American Fisheries Society Symposium*, 23:77–87, 1999.

J. Cadenat. Note sur les merlus de la côte occidentale d'Afrique. In *Congr. Pêche Pêcheries Un. Franç.*, pages 128–130, 1950.

S.X. Cadrin and K.D. Friedland. The utility of image processing techniques for morphometric analysis and stock identification. *Fish. Res.*, 43:129–139, 1999.

S.E. Campana. Chemistry and composition of fish otoliths, pathways, mechanisms and application. *Marine Ecology-Progress Series*, 188:263–297, 1999.

S.E. Campana. Accuracy, precission and quality control in age determination, including a review of the use and abuse of ageing validation methods. *Journal of Fish Biology*, 59: 197–242, 2001.

S.E. Campana. Otolih science entering the 21st century. *Marine and Freshwater research*, 56 (Otolith Research And Applications - Special Issue)(56):485–495, 2005.

S.E. Campana and J.M. Casselman. Stock discrimination using otholit shape analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 50:1062–1083, 1993.

S.E. Campana and C.M. Jones. Radiocarbon from nuclear testing applied to age validation of black drum, pogonias cromis. *Fishery Bulletin U.S.*, 96:185–192, 1998.

S.E. Campana and J.D. Neilson. Microstructures of fish otoliths. *Canadian Journal of Fisheries and Aquatic Sciences*, 42:1014–1032, 1985.

S.E. Campana and S.R. Thorrold. Increments and elements: Keys to a comprehensive understanding of fish populations? *Canadian Journal of Fisheries and Aquatic Sciences*, 58: 30–38, 2001.

S.E: Campana, M.C. Annand, and J.I. McMillan. Graphical and statistical methods for determining the consistency of age determinations. 36:141–151, 1979.

S.E. Campana, K.C.T. Zwanemburg, and J.N. Smith. Pb-210/ra-226 determination of longevity in redfish. *Canadian Journal of Fisheries and Aquatic Sciences*, 47:163–165, 1990.

F. Cao and R. Fablet. Automatic morphological detection of otolith nucleus. *Pattern Recognition letters*, 27:658–666, 2006.

M. Cardinale, P. Doering-Arjes, M. Kastowsky, and H. Mosegaard. Effects of sex, stock, and environment on the shape of known-age Atlantic cod (*gadus morhua*) otoliths. *Canadian Journal of Fisheries and Aquatic Sciences*, 61:158–167, 2004.

J.K. Carlson, E. Cortés, and A.G. Johnson. Age and growth of the blacknose shark, *carcharhinus acronotus*, in the eastern gulf of Mexico. *Copeia*, 1999:684–691, 1999.

S.J. Carpenter, J.M Erickson, and F.D. Holland Jr. Migration of late cretaceous fish. *Nature*, 423:70–74, 2003.

V. Caselles, T. Coll, and J. Morel. Topographic maps and local contrast changes in natural images. *Int. J. of Comp. Vision*, 33(1):5–27, 1999.

M.F. Castelnau. *Merluccius capensis* (n. sp). In *Mémoire sur les poissons de l'Afrique australe*, pages 68–69, 1861.

M. Castonguay, P. Simard, and P. Gagnon. Usefulness of Fourier analysis of otolith shape for Atlantic mackerel (*scomber scombrus*) stock discrimination. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(48):296–302, 1991.

W. Y. B. Chang. A statistical method for evaluating the reproducibility of age determination. *Canadian Journal of Fisheries and Aquatic Sciences*, 39:1208–1210, 1982.

Y. Chen, D.A. Jackson, and H.H. Harvey. A comparison of von Bertalanffy and polynomial functions in modelling fish growth data. *Can. J. Fish. Aquat. Sci.*, 49:1228–1235, 1992.

D. Chesmore. The Automated Identification of Taxa: Concepts and applications. In CRC Press N. MacLeod, editor, *Identification is Systematics: Theory, approaches and applications*, pages 83–100. Academic Press, 2007.

NC. Chui. *Wavelets: a tutorial in theory and applications*. Academic Press, 1992.

M. Cocchi, R. Seeber, and A. Ulrici. Wpter: Wavelet-packet transform for efficient pattern recognition signals. *Chemometrics and Intelligent Laboratory Systems*, 57:97–119, 2001.

A. Cohen, I. Daubechies, and J.C. Feauveau. Biorthogonal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, 45(5):485–560, 1992.

H. Cohen. *A course in computational algebraic number theory*. Springer Verlag, New York, 1993.

R.R. Coifman and F. Majid. Adapted waveform analysis and denoising. In *Progress in Wavelet Analysis and Applications*, pages 63–76. Editions Frontieres, BB.P.33, 91192 Gif-sur-Yvette Cedex, France, 1993.

R.R. Coifman and Y. Meyer. Nouvelles bases orthonormées de l2(r) la structure de systèm de walsh. Technical report, Yale University, 1989a.

R.R. Coifman and Y. Meyer. Orthonormal wave packet bases. Technical report, Yale University (in ftp pascal.math.yale.edu/pub/wavelets/wavepkt.tex), 1989b.

R.R. Coifman and Y. Meyer. Remarques sur l'analyse de fourier a fenêtre. *C.R. Acad. Sci.*, pages 54–81, 1991.

R.R. Coifman and M.V. Wickerhauser. Entropy-based algorithms for Best-Basis selection. *IEEE Transactions on Information Theory*, 2:713–719, 1992.

J.W. Cooley and J.W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comp.*, 19:297–301, 1965.

J.M. Cope and A.E. Punt. Admitting ageing error when fittng growth curves: an example using the von Betarlanffy growth function with random effects. *Can. J. Fish. Aquat. Sci.*, 64:205–218, 2007.

P.E. Cottrell, A.W. Trites, and E.H. Miller. Assessing the use of hard parts in faces to identify harbour seal prey: results of captive feeding trials. *Canadian Journal of Zoology-Revue Canadien de Zoologie*, 74:875–880, 1996.

T.M. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13:21–27, 1967.

I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics*, 41:909–996, 1988.

I. Daubechies. Ten lectures on wavelets. In *CBMS-NSF Regional Confference Series in Applied Mathematics*, volume 61, 1992.

G.F. de Montricher, R.A. Tapia, and J.R. Thompson. Nonparametric maximum likelihood estimation of probability densities by penalty function methods. *Ann. Statist.*, 3:1329–1348, 1975.

P.A. Devijer and J. Kittler. *Pattern recognition: A statistical approach.* Prentice Hall International Inc, London, 1982.

D.A. DeVries, C.B. Grimes, and M.H. Prager. Using otolith shape anlysis to distinguish eastern gulf of Mexico and Atlantic Ocean stocks of king mackerel. *Fisheries Research*, 57:51–62, 2002.

J.D. Dutil, M. Castonguay, D. Gilbert, and D. Gascon. Growth, condition and environmental relationships in Atlantic cod (*gadus morhua*) in the northen Golf of st. Lawrence and implications for management strategies in the North-west Atlantic. *Canadian Journal of Marine Science*, 60:1123–1138, 1999.

B. Efron. Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, 78:316–331, 1983.

R. Fablet. Statistical learning applied to computer-assisted fish age and growth estimation from otolith images. *Fisheries Research*, 81(81):219–228, 2006.

R. Fablet and N. Le Josse. Automated fish age estimation from otolith image anlysis. *Fisheries Research*, 72:279–290, 2005.

R. Fablet, A. Benzinou, and C. Doncarli. Robust time-frequency estimation in otolith images for fish age and growth analysis. In *Proc. of 10th Int. Conf. on Image Processing*, 2003.

A.G. Finstand. Growth backcalculation based on otoliths incorporataing an age effect. *Journal of Fish Biology*, 62:1222–1225, 2003.

R.A. Fisher. The use of multiple measures in taxonomic problems. *Annals of Eugenics*, 7: 179–188, 1936.

J.E. Fitch and R.J. Lavenberg. *Marine food and game fishes of California.* The University of Chicago Press, 1971.

E. Fix and J. Hodges. Discriminatory analysis, nonparametric discrimination: consistency properties. Technical report, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

R. Fletcher. *Pratical methods of optimization (2nd. Edition)*. John Wiley and Sons, New York, p113, 1987.

P. Franca. Nova contribuiçao para o conhecimento do género *Merluccius* no Atlântico oriental ao sul do Equador. *Mem. Junta Invest. Ultram*, 2(18):57–101, 1960.

J.H. Freedman and P. Diaconis. On the histogram as a density estimator: L2 theory. *Zeitschrift für Wahrscheinlichkeitssheorie und verwandte Gebiete*, 57:453–476, 1981.

K.S. Fu. *Syntactic Pattern recognition*. Prentice Hall, Inc, N.J., 1982.

K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, San Diego, CA, 1990.

L.P. Garrison and J.S. Link. Diet of five hake species in the North-east United States continental shelf ecosystem. *Mar. Ecol. Prog. Ser.*, 204:243–255, 2000.

R.W. Gauldie. The morphological basis of fish age estimation methods based on the otolith of *nemadactylus macropterus*. *Canadian Journal of Fisheries and Aquatic Sciences*, 51: 2341–2362, 1994.

A.J. Geffen. Otolith ring deposition in relation to growth rate in herring (*clupea harengus*) and turbot (*scophtkalmus maximus*) larvae. *Marine Biology*, 71:317–326, 1982.

I.J. Good. The interaction algorithm and practical Fourier analysis. *J.R. Statist. Soc. B*, 20(2):361–372, 1958.

I.J. Good and R.A. Gaskings. Density estimation and bump-hunting by the penalized likelihood method exemplified by the scattering and meteorite data (with discussion). *J. Amer. Statist. Assoc.*, 75:42–73, 1980.

C. Goutte. Note on free lunches and cross-validation. *Neural Computation*, 9:1211–1215, 1997.

A. Grant. Age determination of growth in fish and other aquatic animals. *Australian Journal of Marine and Freshwater Research*, 43 (Special issue):879–1330, 1992.

A. Guichenot. *Fauna Chilena. Peces*. C. Gay, 1848.

A. Guillaud, A. Benzinou, H. Troadec, V. Rodin, and J.L. Bihan. Autonomous agents for edge detection and continuity perception on otolith images. *Image and Vision Computing*, 20:955–968, 2002.

L. Han and M. Neumann. Effect of dimensionality on the Nelder-Mead simplex method. *Optimization methods and software*, 21(1):1–16, 2006.

S.M. Hanchet and Y. Uozumi. Age validation and growth of southern blue whiting, *micromesitius australis*, in New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 30:57–67, 1996.

D.J. Hand. *Kernel discriminant analysis*. Chichester: Research Studies Press, 1982.

E. Hecht. *Optics, 3rd. edition*. Pearson Education, 2000.

C. Herley and M. Vetterli. Linear phase wavelets: Theory and design. In *Proc. IEEE Int. Conf. ASSP (Toronto, Canada)*, pages 11–21, 1991.

C. Herley and M. Vetterli. Wavelets fractals and fourier transforms. In J.C.R. Hunt M. farge and Eds. J.C. Vassilicos, editors, *Biorthogonal bases of symmetric compactly supported wavelets*, pages 289–297. Oxford University Press, 1992.

M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. A real-time algorithm for signal analysis with the help of the Wavelet transform. In A. Grossmann and Ph. Tchamitchian, editors, *Wavelets, Time-frequency methods and Phase space*, pages 289–297. Springer-Verlag, Berlin, 1989.

F.W. Hutton. Fishes of New Zealand. Catalogue with diagnoses of the species. Technical report, Colonial Museum and Geological Survey Dep., 1872.

G.A. Hyndres, N.R: Loneragan, and I.C. Potter. Influence of sectioning otoliths on marginal increment trends and age and growth estimates for the flathead, *platycephalus speculator*. *Fisheries Bulletin U.S.*, 90:276–284, 1992.

T. Inada. Studies on the merluccid fishes. *Bulletion of Far Seas Reserach laboratory (Shimizu)*, 18:1–172, 1981.

R. Jain, R. Kasturi, and B.G. Schunck. *Machine Vision*. Mcgraw - Hill, 1995.

I.T. Jollife. *Principal Component Analysis*. Springer, 1987.

M.C. Jones. Variable kernel density estimates. *Austral. J. Statist*, 32:361–371, 1990.

T. Kailath. *Linear Systems*. Englewood Cliffs, NJ: Prentice Hall, 1980.

J.M. Kalish. Determinants of otolith chemistry: seasonal variation in the composition of blood plasma, endolymph and otolith beareded rock cod. *pseudophycis barbatus. Marine Ecology-Progress Series*, 74:137–159, 1991.

J.M. Kalish. Pre- and post-bomb radiocarbon in fish otoliths. *Earth and Planetary Science Letters*, 114:549–554, 1993.

J.M. Kalish. Application of the bom radiocarbon chronometer to the validation of redfish *centroberyx affinis* age. *Canadian Journal of Fihseries and Aquatic Sciences*, 52:1399–1405, 1995a.

J.M. Kalish. Radiocarbon and fish biology. In S.E. Campana D.H. Secor, J.M. Dean, editor, *Recent Developments in Fish Otolith Research*, pages 637–653. UColumbia: University of South Carolina Press, 1995b.

C. Karlou-Riga. Otolith morphology and age and growth of *trachurus mediterraneus* in the Eastern Mediterranean. *Fisheries Research*, 46:69–82, 2000.

R. Klette, K. Schlüns, and A. Koschan. *Computer Vision: Three-Dimensional data from Images*. Springer Verlag, 1998.

V.K. Klonias. Consistency of two nonparametric maximun penalized likelihood estimators of the probability density function. *Ann. Statist.*, 10:811–824, 1982.

R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International Joint Conference on Artificial Intelligence*, 1995.

T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola. Lvqpak: A program package for the correct application of Learning Vector Quantization algorithms. In *Proceedings on International Joint Conference on Neural Networks*, 1992.

T. Kohonen, J. Hynninen, J. Kangas, Laaksonen J, and K. Torkkola. Lvqpak: A program package for the correct application of Learning Vector Quantization algorithms, 1995. URL `http: //www.cis.hut.fi/research/lvqpak/`. December 2009.

R.A. Kronmal and M.E. Tarter. The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.*, 63:925–952, 1968.

F.P. Kuhl and C.R. Giardina. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, 18:236–258, 1981.

S. Kullback and R.A Liebler. On information and sufficiency. *Annals of Mathematical and Statistics*, 22:79–86, 1951.

P.A. Lachenbruch. *Estimation of error rates in discriminant analysis*. PhD thesis, University of California, Los Angeles, 1965.

F. Lagardère and H. Troadec. Age estimation in common sole (*solea solea*) larvae: validation of daily increments and evaluation of a pattern recognition technique. *Marine Ecology-Progress Series*, 155:223–237, 1997.

T.E. Laidig, D.E: Pearson, and L.L. Sinclair. Age and growth of blue rockfish (sebastes mistinus) from Central and Northern California. *Fishery Bulletin*, 101:800–808, 2003.

A.H. Land and A.G. Doig. An automatic method of solving discrete programming problems. *Econometrica*, 28:497–520, 1960.

K. Levenberg. A method for the solution of certain problems in Least-Squares. *The Quaterly of Applied Mathematics*, 2:164–168, 1944.

C. Linnareus. *Systema Naturae*, volume 1. –, 1758.

D. Lloris and J. Matallanas. Description of a new species of hake: *merluccius patagonicus* sp. nov. (gadiformes, merluciidae) from the waters of argentina. *Scientia Marina*, 67:323–326, 2003.

D. Lloris, J. Matallanas, and P. Oliver. *Hakes of the world (Family Merluciidae). An annotated and ilustrated catalogue of Hake species known to date*. FAO, 2005.

A. Lombarte. *Efectos ambientales y filogenéticos en la morfología del laberinto audutivo de los peces Teleósteos*. PhD thesis, Universitat de Barcelona, 1990.

A. Lombarte and A. Castellón. Interspecific and intraspecific otolith variability in the genus merluccius as determined by image analysis. *Canadian Journal of Zoology-Revue Canadien de Zoologie*, 69(69):2442–2449, 1991.

A. Lombarte and J.M. Fortuño. Diferences in morphological features of the *sacculus* of the inner ear of two hakes (*Merluccius capensis* and *Merluccius paradoxus*, gadiformes) inhabits from different depth of sea. *J. Morphol.*, 214:97–107, 1992.

A. Lombarte and J. Lleonart. Interspecific and intra-specific otolith variability in the genus *Merluccius* as determined by image analysis. *Environmental Biology of Fishes*, 37:297–306, 1993.

A. Lombarte, I. Olaso, and A. Bozzano. Ecomorphological trends in artedidraconidae (Pisces: *Pesciformes Notothenioidei*) of the Weddell Sea. *Antartic Science*, 15, 2003.

A. Lombarte, Ò. Chic, V. Parisi-Baradad, R. Olivella, J. Piera, and E. Garcia-Ladona. A web-based enviroment from shape analysis of fish otoliths: The AFORO database. *Scientia Marina*, 70:147–152, 2006.

N. MacLeod. *Automated Taxon Identification is Systematics: Theory, approaches and applications*. CRC Press, 2007.

S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Learning*, 11:674–693, 1989.

S.G. Mallat. *A wavelet tour of signal processing*. Academic Press, 1990.

S.G. Mallat and W.L. Hwang. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 32:617–643, 1992.

B. Marchand and N. Saito. Earth Mover's Distance-based Local Discriminant Bases, 2012. URL `www.math.ucdavis.edu`. Springer to appear.

D. Marquardt. An algorithm for Least-Squares estimation of non-linear parameters. *SIAM journal of Applied Mathematics*, 11:977–992, 1963.

W.J. McCurdy, J. Panfili, E.J. Meunier, A.J. Geffen, and H. de Pontual. Preparation of calcified structures. In H Troadec P.J. Wright J. Panfili, H de Pontual, editor, *Manual of fish Sclerochronology*, pages 331–357. Brest: France, Ifremer-IRD Coedition, 2002.

G.J. McLachlan. *Discriminant analysis and statistical patern recognition*. Wiley-Interscience paperback series. J. Wiley and Sons, New York, 1992.

S.N. Messieh, C. McDougall, and R. Claytor. Separation of Atlantic herring (*clupea harengus*) stocks in the Southern Gula of St. Lawrence using digitized otolith morphometrics and discriminant function analysis. *Canadian Technical Report of Fisheries and Aquatic Sciences*, 1647:1–22, 1989.

Y. Meyer. *Wavelet and operators*. Cambridge Univ. Press, New York, 1990.

Y. Meyer. Wavelets: Algorithms and applications. Technical report, SIAM, 1993.

F. Mintzer. Filters for distortion-free two-band multirate filter banks. *IEEE. Trans. Acoust., Speech, Signal Processing*, 33:626–430, 1985.

J. Mitchill. Description of three species of fish. *J. Acad. Nat. Sci. Phila*, 1:407–412, 1818.

S.L. Mitchill. Report in part on the fishes of New York. *Trans. Lit. Phil. Soc.*, 1:7, 1814.

L.R. Mombeck. Notes on the distinction of the North-west Atlantic hakes, *merluccius albidus* and *merluccius bilinearis*. *International Convention of the Northwest Atlantic Fisheries*, 8:87–89, 1970.

P. Monasse and F. Guichard. Fast computation of a contrast invariant representation. *IEEE Transactions on Image Processing*, 9(5):860–872, 2000.

B. Morales-Nin. Growth determination of tropical marine fishes by means of otolith interpretation and length frequency analysis. *Aquatic Living Resources*, 2:241–253, 1989.

B. Morales-Nin, G.J.Torres, A. lombarte, and L. Recasens. Otolith growth and age estimation on european hake *merluccius merluccius*. *Journal of Fish Biology*, 53:1155–1168, 1998a.

B. Morales-Nin, A. Lombarte, and B. Japón. Approaches to otolith age determination: image, signal treatement and age atribution. *Scientia Marina*, 62:247–256, 1998b.

A.K. Morison, S. Robertson, and D. Smith. An integrated system for production fish aging: image analysis and quality insurance. *North American Journal of Fisheries Management*, 18:587–598, 1998.

A.K. Morison, J. Burnett, W.J. McCurdy, and E. Mokness. Quality issues in the use of otoliths for fish age estimation. *Marine and Freshwater research*, 56 (Special Issue: Fish Otolith Research and Applications):773–782, 2005.

K. Morita and T. Matsuishi. A new model of growth-back calculation incorporating age effect based on otoliths. *Canadian Journal of Fisheries and Aquatic Sciences*, 60:1376–1385, 2001.

H. Mosegaard, B. Morales-Nin, C. Weidman, A. Geffen, E. Ameri, R. Millner, J. Panfili, and A. Folkvord. Otolith thin-section preparation: some problems and new developments. In P.J. Wright, editor, *Otolith preparation and analysis. Proceedings of European Fish Ageing Network Workshop*, volume 2, pages 11–21, 1998.

L.J. Natanson, J.J. Mello, and S.E. Campana. Validate age and growth of the porbeagle shark, *lamna nasus*, in the Western North Atlantic Ocean. *Fishery Bulletin U.S.*, 100: 266–278, 2002.

J.A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7(4):308–313, 1965.

D. Nolf. Otolith piscium. In H.P. Schultze, editor, *Handbook of paleoichthyology*, volume 10, pages 1–145. Gustav Fisher Verlag, 1985.

J.R. Norman and P.H. Greenwood. *A history of fishes*. Ernst Bennt Lmtd., 1975.

K. Oliveira. Field validation of annular growth rings in the american eel, *anguilla rostrata*, using tetracycline-marked otoliths. *Fishery Bulletin U.S.*, 94:186–189, 1996.

R.G. Osuna. Prism: Pattern recognition and intelligent sensor systems. web, 2005. URL `http://research.cs.tamu.edu/prism/lectures/`. Accessed on october 2011.

N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans*, 9(1):62–66, 1979.

J. Panfili, H. de Pontual, H. Troadec, and P.J. Wright. *Manual of Fish Sclerochronology*. Brest: France, Ifremer-IRD Coedition, 2002.

V. Parisi-Baradad, A. Lombarte, E. Garcia-Ladona, J. Cabestany, J. Piera, and O Chic. Otolith shape contour analysis using affine transformation invariant Wavelet transforms and Curvature scale space representation. *Marine and Freshwater research*, 56(56):795–804, 2005.

V. Parisi-Baradad, A. Manjabacas, A. Lombarte, R. Olivella, Ò. Chic, J. Piera, and E. García-Ladona. Automated taxon identification of teleost fishes using and otolith online database - AFORO. *Fish. Res.*, 2010. doi: 10.1016/j.fishres.2010.02.005.

B.B. Parrish and D.P. Sharman. Otolith tupes among summer-autumm spawning herring in the Northern North Sea. *J. Cons. Int. Expl. Merc.*, 25:81–92, 1959.

T. Pavlidis. *Structural Pattern recognition*. Springer Verlag, New York, 1977.

J. Piera, V. Parisi, E. García-Ladona, A. Lombarte, L. Recasens, and J. Cabestany. Otolith shape feature extraction oriented to automatic classification with open distributed data. *Marine Freshwater Research*, 56:805–814, 2005.

G.M. Pilling, G.P. Kirkwood, and S.G. Walker. An improved method for estimating individual growth variability in fish, and the correlation between bertalanffy growth parameters. *Can. J. Fish. Aquat. Sci.*, 59(3):424–432, 2002.

K.R. Piner, O.S. Hamel, J.L. Menkel, J.R. Wallace, and C.E. Hutchinson. Age validation of canary rockfish (*sebastes pinniger*) from off the Oregon Coast (USA) using the Bomb-radiocarbon method. *Can. J. Fish. Aquat. Sci.*, 62(5):1060–1066, 2005.

T.J. Pitcher and P.D.M. McDonald. Two models for seasonal growth in fishes. *J. Appl. Ecol.*, 10(2):599–606, 1973.

M. Plancherel and M. Leffler. Contribution à l'étude de la représentation d'une fonction arbitraire par les intégrales définies. *Rendiconti del Circolo Matematico di Palermo*, 30: 289–235, 1910.

C. Platt and A.N. Popper. Fine structure and function of the ear. In A.N. Popper W.N. Tavolga and R.N. Fay, editors, *Hearing and communication in fishes*, pages 3–36. Springer-Verlag, 1981.

N. Poulet, P. Berrebi, A.J. Crivelli, S. Lek, and C. Arguillier. Genetic and morphometric variations in the pikeperch (*sander lucioperca* l.) of a fragmented delta. *Archiv Fur Hidrobiologie*, 159:531–554, 2004.

Prajneshu and R. Venugopalan. von Bertalanffy growth model in a random environment. *Can. J. Fish. Aquat. Sci.*, 56:1026–1030, 1999.

J.G. Proakis and D.G. Manolakis. Discrete-time signals and systems. In International Inc. Prentice Hall, editor, *Digital Signal Processing: Principles, algorithms and applications*, pages 22–27. Prentice Hall, International Inc., 1996.

T.J. Quinn and R.B. Delriso. *Quantitative fish dynamics*. Oxford University Press, New York, 1999.

T.P. Quinn, R.S. Nemeth, and D.O. McIsaac. Homing and straying patterns of fall chinook salmon in the Lower Columbian River. *Transactions of the American Fisheries Society*, 120:150–156, 1991.

L.R. Rabiner and R.W. Schafer. Inroduction to digital signal processing. *Foundations and Trends in Signal Processing*, 1(1-2):1–194, 2007.

K. R. Rao and P. Yip. *Discrete Cosine transforms: Algorithms, advantages, applications*. Academic Press, 1990.

S. Reeves. A simulation of the implications of age-reading errors for stock assessment and management advice. *ICES Journal or Marine Science*, 60:314–328, 2003.

O. Rioul. Regular wavelets: a discrete-time approach. *IEEE Trans. Signal Processing*, 41: 3572–3579, 1993.

J. Rissanen. A universal prior for integers and estimation by Minimum Description Length. *Ann. Statist.*, 11(2):416–431, 1983.

F.J. Rohlf and L.F. Marcus. A revolution in morphometrics. *Trends in Ecology and Evolution*, 8:129–132, 1993.

M.I. Roldan, J.L. Garcia-Marin, F.N. Utter, and C. Pla. Genetic relations among *Merluccius* species. *Heredity*, 83:79–86, 1999.

F. Rosemblatt. The perceptron: A perceiving and recognizing automaton. Technical report, Cornell Aeronautical Laboratory, 1957.

M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9:65–78, 1982.

N. Saito. *Local Feature Extraction and its applications using a library of bases.* PhD thesis, Yale University. Department of Mathematics, 1994. Available at: http://www.math.ucdavis.edu.

N. Saito and RR. Coifman. Local Discriminant Bases and their applications. *J. Math. Imaging Vis.*, 5:337–358, 1995.

N. Saito and E. Woei. Simultaneous segmentation, compression and denoising of signals using polyharmonic Local Sine transformation and Minimum Description Length criterion. In *13th. IEEE Statistical signal processing workshop*, pages 315–320, 2005.

N. Saito, RR. Coifman, F.B. Geshwind, and F.Warner. Discriminant feature extraction using empirical probability density estimation and a local basis library. *Pattern Recognition*, 35: 2841–2852, 2002.

J.W. Sammon. A non-linear mapping for data structure analysis. *IEEE Transactions on Computers*, 18(5):401–409, 1969.

M. R. Schroeder. *Number theory in Science and Communications 2nd. Ed.* New York: Springer Verlag, 1986.

D.W. Scott. On optimal and data-based histograms. *Biometrika*, 66:605–610, 1979.

D.W. Scott. Frequency poligons. *Journal of the American Statistical Association*, 80:348–354, 1985a.

D.W. Scott. Averaged Shifted Histograms: Effective nonparametric density estimation. *Annals of Statistics*, 13:1024–1040, 1985b.

D.W. Scott. *Multivariate density estimation.* Jon Willey & Sons, 1992.

D.W. Scott and G.R. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82:1131–1146, 1987.

D.H. Secor, , J.M. Dean, and S.E. Campana. *Recent Developments in Fish Otolith Research.* University of South Carolina, 1995a.

D.H. Secor, E.D: Houde, and D.M. Monteleone. A mark-release experiment on larval stripped bass *Morone saxatilis* in a Chesapeak bay tributary. *ICES Journal of Marine Science*, 52: 87–101, 1995b.

C.A. Silva-Segundo, M. Brito-Chavarria, E.F. Balart, I.A. Barriga-Sosa, R. Rojas-Equivel, M.I. Roldán, G. Murugan, and J. Garcia de León. Clarifying taxonomic status of merluccius spp. in the northeastern pacific: a combined morphological and molecular approach. *Rev. Biol. Fish. Fisheries*, 21:259–282, 2011.

B.W. Silverman. *Density estimation for Statistics and Data analysis.* Ckapman and Hall, 1986.

N.V. Slyke. A review of the analysis of fish remains in Chumash sites. *Pacific Coast Archeological Society Quaterly*, 34:25–58, 1998.

M.J. Smale, G. Watson, and T. Hecht. Otolith atlas of Southern African Marine Fishes. *Ichthyological Monographs of the J.B.L. Smith Institute of Ichthyology*, 1:1–232, 1995.

C. Smith. Some examples of discrimination. *Ann. Eugen.*, 18:272–282, 1947.

M.J.T. Smith and T.P. Barnwell. A procedure for designing exact reconstruction filter banks for tree structres subband coders. In *Proc. IEEE Int. Conf. ASSP, San Diego, Mar. 1984*, pages 27.1.1–27.1.4, 1984.

J.A. Soria and V. Parisi-Baradad. Best-Basis development towards the automatical detection of otolith irregularities in fishes. In *7th International Conference on Next generation Web Services and Practices*, pages 332–337, 2011. doi: 10.1109/NWeSP.2011.6088200.

J.A. Soria, A. Lombarte, and V. Parisi. Local Discriminant Bases representation and non-linear growth processing for species classification and age estimation of fish based on otolith images. In *MTS/IEEE Kobe Techno-Ocean 2008*, volume 1-3, pages 312–317, 2008.

J.A. Soria, A. Lombarte, and V. Parisi-Baradad. Otolith identification of *Merluccius* populations and sympatric species with Local Discriminant Bases. In *4th. International Symposium on fish otolith research and applications, Monterey-EEUU*, 2009.

W. Spendley, G.R. Hext, and F.R. Himsworth. Sequential application of simplex designs optimisation and evolutionary operation. *Technometrics*, 4(4):441–461, 1962.

C.J. Stone. Consistent non-parametric regression (with discussion). *Annals of Statistics*, 5: 595–645, 1974.

G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wesley-Cambridge Press, 1996.

H.A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21:65–66, 1926.

W. Sweldens. The lifting scheme: a custom-designed construction of biorthogonal wavelets. *J. of Appl. and Comput. Harmoniuc Analysis*, 3(2):186–200, 1996.

W. Sweldens. The lifting scheme: a construction of second generation wavelets. *SIAM J. of Math. Analysis*, 29(2):511–546, 1997.

G.R. Terrel and D.W. Scott. Oversmoothed nonparametric density estimation. *J. Amer. Statist. Assoc.*, 80:209–214, 1985.

D'A. W. Thompson. *On growth and Form*. Cambridge University Press, 1917.

W.M. Thorburn. Occam's razor. *Mind*, 24:287–288, 1915.

G.J. Torres. *Variación intraespecífica e interespecífica de la morfología y de las pautas del crecimiento del otolito sagitta en el género Merluccius (Gadiformes: Teleosti)*. PhD thesis, Polytechnic University of Catalonia, 1997.

G.J. Torres, A. Lombarte, and B. Morales-Nin. Sagittal otolith size and shape variability to identify geographical intra-specific differences in three species of the genus *Merluccius*. *Journal of the Marine Biological Association of the United Kingdom*, 80:333–342, 2000a.

G.J. Torres, A. Lombarte, and B. Morales-Nin. Variability of the *sulcus acusticus* in the *sagitta* otolith of the genus *Merluccius*. *Fisheries Research*, 46:5–13, 2000b.

G.T. Toussaint. Bibliography on estimation of misclassification. *IEEE Transactions on Information Theory*, 20:472–479, 1974.

H. Troadec. Frequency demodulation on otolith numerical images for the automation of fish age estimation. *Aquatic Living Resources*, 4:207–219, 1991.

H. Troadec, A. Benzinou, V. Rodin, and J.L. Bihan. Use of deformable templates for otolith 2D growth ring detection by digital image processing. *Fisheries Research*, 46:155–163, 2000.

V.A. Tuset, A. Lobarte, and A. Assis. *Otolith atlas for the Western Mediterranean, North and Central Eastern Atlantic*. Canada Communication Group, 2008.

V.M. Tuset, S. Piretti, A. Lombarte, and J.A. Gonzalez. Using otoliths and eye diameter for ecological characterization of depp-sea fish: *Aphonopus carbo* and *A. intermedius* from NE Atlantic waters. *Scientia Marina*, 74(4):807–814, 2010. doi: 10.3989/scimar.2010.74n4807.

K. Tuskamoto and T. Kajihara. Age determination of ayu with otolith. *Nippon Suisan Gakkaishi*, 53:1985–1997, 1987.

P.P Vaidyanathan. Multirate digital filters, filter banks, polyphase networks and applications. *Proc. IEEE*, 78:56–93, 1990.

P.P Vaidyanathan and P.Q. Hoang. Lattice structures for optimal design and robust implementation of two-chanel perfect reconstruction qmf filter banks. *IEEE Trans. Acoust. Speech Signal Process.*, 36:81–94, 1988.

V. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, 1995.

M. Vetterli. Filter banks allowing perfect reconstruction. *Signal Processing*, 37:1057–1071, 1986.

M. Vetterli. Wavelets and filter banks: Theory and design. *IEEE Transactions on Signal Processing*, 40(9):2207–2231, 1992.

L. Vincent. Grayscale area openings and closings, their efficient implementation and applications. In J. Serra and P. Salembier, editors, *Proc. of the 1st. Workshop on Math. Morphology and its App. to Sig. Proc.*, pages 22–27. 1993.

Y.-G. Wang and M.R: Thomas. Accounting for individual variability in the von Bertalanffy growth model. *Can. J. Fish. Aquat. Sci.*, 52:1368–1375, 1995.

S. Watanabe. Karhunen-loève expansion and factor analysis: theoretical remaks and applications. In *Trans. 4th Prague Conf. Inform. Theory, Statist. Decision Functions , Random Processes*, pages 635–660. Publishing House of the Czechoslovak Academy of Sciences, 1967.

S. Watanabe and T. Kaminuma. Recent developments of the minimum entropy algorithm. In *Proceedings of the International Conference on Pattern Recognition*, pages 536–540, 1988.

G.S. Watson. Density estimation by orthogonal series. *Ann. Math. Statist.*, 40:1496–1498, 1969.

J.E. Watson. A technique for mounting and storing herring otoliths. *Trans. Ann. Fish. Soc.*, 94:267–268, 1965.

H. Welleman and F. Storbeck. Automatic ageing of plaice otoliths by means of image analysis. In D. Dean and S.E. Campana, editors, *Recent developments in fish otolith research*, pages 272–282. University of South Carolina Press, 1995.

J.A. Wheeler and H. Zureck. *Quantum theory and measurement*. Univ. Press, 1983.

M.V. Wickerhauser. *Adapted wavelet analysis: From theory to software*. Wellesley, 1995.

C.A. Wilson and D.L. Neiland. Age and growth of red snapper (*lutjanus campechamus*) from the Northern Gulf of Mexico of Louisiana. *Fishery Bulletin*, 99:653–664, 2001.

D.T. Wilson and M.I. McCormick. Spatial and temporal validation of settlement marks in the otoliths of red fishes. *Marine Ecology Progress Series*, 153:259–271, 1997.

P.J. Wright, J. Panfili, B. Morales-Ni, and A.J. Geffen. Otoliths. In J. Panfili, H de Pontual, H. Troadec, and P.J. Wright, editors, *Manual of fish sclerochronology*, pages 31–57. Ifremer-IRD coedition, 2002.

Subject Index

accuracy validation
- bootstrap, 85
- hold-out(split-sample), 83–84, 127, 153
- K-fold cross-validation, 84, 104, 130, 150, 153, 163
- leave-one-out, 84
- re-substitution, 83

age estimation
- advanced tools
  - Neural networks, NN, 6, 9, 95
  - Peak-based rprst., PB, 37–38
- age structures, 6, 10–23, 30–38, 111–114, 148–151, 168, 169
- average percentage error(APE), 21
- coefficient variation(CV), 22, 33–35
- conventional tools
  - Adapted-growth filt., AGF, 36
  - LDA-based classifier, 87
  - Levengberg-Marquardt algorithm, LMA, 33–36
  - Manual contrast extraction, MCE, 36
- expert knowledge, 7–8, 32, 103, 147–151
- gray level, 6, 15–39, 159
- growth estimation, 6, 11, 20–39, 100, 111–118, 147–170
- image contrast, 6, 12, 19–26, 36–39, 112–120, 150–152, 165–170
- intensity profile, 6, 12, 25–46, 100, 112–120, 147–159, 168–169
- inter-expert agreement, 22, 150
- manual age validation, 20–22
- nucleus detection, 26–30, 38
- radial, 6, 31–33, 39, 112, 147
- ring, 6–38

automatic, 5, 7–10, 15, 25, 39, 43, 70, 99, 105, 111, 114, 118, 132, 147, 154, 165, 166, 169

basis function
- bi-orthogonal basis, 56, 123, 186–189, 192–196
- compact support, 52, 153, 161, 177–180, 196
- orthogonality, 52–56, 66–72, 173–191
- orthonormality, 49–76, 109–111, 175–184
- perfect reconstruction, 52–60, 176–193

Best-Basis paradigm
- conventional component analysis, 9, 44, 70–73, 79, 121–123
- definition, 8, 43
- Earth mover's distance, EMD, 10
- Karh.-Loeve Basis, KLB, 70–71

Local Discriminant Basis, LDB, 11–13, 46, 73–76, 79, 99
Thesis proposition
Density Local Discriminant Bases, DLDB, 107–111, 123–151

classification, 1–12, 36–46, 71–170
classifiers
Bayes rule, 9, 85–87, 93, 121
calibration set, 83–88, 94, 163
Learning Vector Quantization(LVQ), 10, 89–94
linear
k-Nearest Neighbor(k-NN), 9, 88–94
LDA-based, 73, 81–88
misclassification rate, 82–90
partition, 82–89, 108, 163
supervised learning, 13, 81
testing set, 83–90, 127
validation of accuracy, 82–85
component analysis
between-class variance, 72, 145
covariance matrix, 68–73, 87, 145
discriminant components, LDA, 71–73, 79, 121–123
eigenvalues, 71–72, 127, 157
eigenvectors, 70–72, 145
Euclidean distance, 28, 78, 88
Fisher's separability index, 72–73
principal components, PCA, 9, 44, 68–71, 79, 121–123
within-class variance, 24, 72, 145
contour normalization
elliptical Fourier series, 101, 120, 153, 160
homology, 153–159
radial, 101–103
rotation, 101–103, 114, 160
size, 101
translation invariance, 11, 38, 56, 67, 100–101, 153–160

data preprocessing
contour normalization, 101–103
contrast cancellation, 36–38, 120, 150–152, 165–168
digitization, 46
from 2D to 1D, 31
segmentation, 15, 23–25, 38
signal demodulation, 12–13, 32–39, 100, 111–117, 148–158, 168

experiments and results
algorithms
AGD (example), 147–148
DFT, 121–129
k-NN, 127–129
LDB vs. DLDB, 123–129
LVQ, 127–129
PCA and LDA, 121–129
application to aging
DLDB and LVQ, 149–151
MCE, AGF, SGD and AGD, 148–151
inter-specific studies
M. albidus - *M. bilinearis*, 132–134
M. capensis - *M. paradoxus*, 134
M. gayi - *M. australis*, 137
M. polli - *M. senegalensis*, 134–137
intra-specific studies
M. bilinearis, 139
M. gayi, 139–143
M. merluccius, 138–139
M. productus, 143

validation of fish studies, 130–132

feature extraction, 8–13, 25, 36–79, 95–100, 114–143, 152–170
feature extraction: *concepts*
component analysis, 67–71
data compression, 8–12, 43–48, 62–79, 145, 180
decomposition/analysis, 46–67
density estimation, 9–11, 99–107, 201–207
dimensionality reduction, 8–11, 43, 69–79, 99, 161, 164
discriminant analysis, 71–73
information measures, 8–11, 68–76
reconstruction/synthesis, 46–67
fish ear
epithelium, 2, 3
innervations, 2
labyrinthic cavities, 1
maculae, 2
nerves, 2
otolith, 1–44, 101–102, 111–169
fish identification
contour normalization, 12, 100–103, 114, 139
conventional tools
DFT, 47–48
landmarks, 1
LDA, 71–73
LDA classifier, 87
outlines, 1
PCA, 70–71
inter-specific analysis, 1–3, 11, 44, 118, 130–147, 152–169
morphological studies, 2–5, 44, 101, 139–146
radial length, 25, 101–103, 120–126, 152–159
sympatric studies
geographical relation, 1, 118, 138–143
intra-specific analysis, 1–5, 11, 118, 130–147, 158–169
population dynamics, 3–5, 118, 130, 139–152
fish species
Gadus morhua(cod), 118–127, 143
Merluccius spp.(hake), 2–5, 13, 38, 127–158, 169
Pleuronectes platessa(plaice), 7–16, 36–38, 119, 148–154, 165

growth estimation
curve fitting, 33–36
increment, 20–21, 147
model
Brody coefficient, 33, 148
initial length, 33
maximum length L_∞, 33
model template
Von Bertanlanffy function(VBGF), 11, 33, 112
modulation, 11–13, 100, 111–114, 166
rate, 3, 20, 32, 139–148

input data, 9–10, 54–94, 107, 145–163

Local discriminant Basis, LDB
Bhattacharyya's affinity, 109
bottom-up search strategy, 110, 145
divide-and-conquer, 68–71, 74–76, 100, 110–111, 123, 161

energy map, 73, 109
Kullback-Liebler criterion, 123, 145
optimal trees, 59, 66–71, 100, 108–110, 123, 134, 153
parent nodes, 69, 74, 153
top-down search strategy, 11, 110–111, 145, 157, 161

non-parametric density functions
averaged shifted histogram(ASH), 10, 202–205
bin width, 103–107, 201–207
error estimation
bias and variance, 104–107
ISE, MISE and AMISE, 104–107, 201–204
kurtosis and skewness factor, 105
roughness, 104, 205
FP-ASH, 204
frequency polygons (FP), 202
histogram, 24, 103–112, 130, 148–205
k-Nearest Neighbor(k-NN), 89, 206–207
kernel density estimation(KDE), 89, 153–161, 202–206
probability distribution, 24–30, 49, 72, 81–89, 204
width selection rules
BCV and UCV, 106–107
oversmoothed width, 107
nucleus detection
cardinality, 29
contrario hypothesis, 29–30
geometrical selection, 28–30
grain filtering, 26
maxima removal, 26
number of false alarms(NFA), 29
principal axis proximity, 26
semi-major axis, 26, 28
topographic map, 26

otoliths
asteriscus, 1
calcified structures, 1, 16–20
contour morphology, 2
fisheries management, 3, 20, 38, 117
lapillus, 2
marine applications
age estimation, 5–22, 38–43, 99, 114–118, 147–166
fish identification, 5–25, 38–46, 99–101, 114–120, 158–159
nomenclature
cauda, 2, 132–146, 158
collicum-ostial, 2
collum, 2
dorso-caudal margin, 2, 120–147
dorso-rostral margin, 2, 126–158
rostrum, 2
sulcus acusticus, 2, 134, 146
ventral, 2, 126, 132–143
nucleus, 15, 25–38, 149
preparations for observation
direct observation, 16
embedding, 16–20
reflected-light, 16–22, 154, 158, 166–167
transmitted-light, 16–26, 38
radial, 6, 25–39, 101–103, 112–127, 147–159
sagittae, 1, 16

pattern recognition, 7–22, 43, 67–79, 93–127, 148–160, 170

pattern recognition: *methods*
classification techniques, 81–95
feature extraction techniques, 43–79
preprocessing techniques, 23–38, 101–103, 111–113
preprocessing, 11–13, 15, 25, 38, 99, 117–120, 148–154, 170
projects
AFISA, 12, 117, 148–158
AFORO3D, 12, 117, 130, 158
IBACS, 6

signal feature description
basis function, 45–79, 103–150, 160–164, 181–186, 197, 207
Best-Basis paradigm, 10–12, 44–46, 68–79, 99, 143, 153–163
block-lapped transforms, 59–67
energy entropy, 10, 43, 68–76, 99, 109, 121, 143, 157
feature coefficients, 46–77, 99–143, 173–196
feature space, 8, 43–95, 108–114, 145, 173–187
filter design, 190–196
finite impulse response filter, FIR, 54, 174–196
Fourier transform
FT and DFT, 6–9, 44–48
STFT, 48–49
frequency band, 56, 125–143
frequency domain, 9–10, 36–79, 120–143, 157
graphical tools
sammon map, 77–79
scatter map, 77, 121–127, 157–161
spectrogram, 121, 125
time/frequency momentum, 76–77
irregularities, 125–146, 152–170, 179
Standard Euclidean Basis, STD, 38, 43–45, 79, 117, 121–123, 127–129, 145
time domain, 9, 32, 47–67, 76–77, 120–146, 157, 198–199
time-frequency plane, 11, 48–52, 73–77, 161–170, 175, 197
trigonometric bases
block cosinus-sinus transform, BCT/BST, 65–67
sine-cosine transforms, DCT,DST, 60–62
wavelet series, WS, 50–59

Vector Quantization(VQ)
blank network, 9, 89–95
codebook vector, 89–94
convergence, 92–94
learning rate, 92–94
LVQ1, LVQ2, LVQ3, 90–92
stopping rule, 94

wavelets
à trous algorithm, 153, 160, 198–199
approximations-details, 46, 54–56, 159
continuous transform, CWT, 50–52
dictionary, 44–46, 58–59, 68–76, 109, 123, 129–146, 150, 164
discrete transform, DWT, 54–56, 59
Discrete-time transform, DTWT, 51
filter design
analysis-synthesis filter, 55, 182–199
coiflets, 123, 160, 192
complementary filter, 184–196
Daubechies, 52, 123, 160, 180, 192

Diophantine equations, 173, 194–196
filter-bank, 54, 56, 173, 187–193
low-pass and high-pass, 54, 174–199
paraunitary, 173, 184–187
poly-phase matrix, 183–193
quadrature mirror filter, QMF, 54, 69–77, 109–111, 164
r-biorthogonal, 123
regularity, 159, 179–180, 187–196
symlets, 123, 150, 160, 192
finer scale sampling, 197–198
library, 45, 70, 76, 79, 123, 149
mother wavelet ψ, 45, 50–54, 173–178, 180–182
multi-resolution theory, 52–56, 159, 173–189
scaling function ϕ, 53–54, 173–179, 190
subspaces(nodes), 52–54, 57, 187
two-scale difference equations, 53–57, 159, 178–187
wavelet packet transform, WPT, 56